SCIENCE AND TECHNOLOGY TEXT MINING: ELECTROCHEMICAL POWER

By

Dr. Ronald N. Kostoff
Office of Naval Research
800 North Quincy Street
Arlington, VA 22217
Phone: 703-696-4198
Fax: 703-696-4274
Internet: Kostofr@Onr.Navy.Mil

Mr. Rene Tshiteya
DDL Omni Engineering, LLC
8260 Greensboro Drive, Suite 600
Mclean, VA 22102

Ms. Kirstin M. Pfeil
Office of Naval Research
Arlington, VA

Mr. James A. Humenik
Noesis, Inc.
Manassas, VA

(*The views in this paper are solely those of the authors, and do not represent the views of the Department of the Navy or any of its components, DDL-OMNI, Inc., or Noesis, Inc. in addition, the authors acknowledge the contributions of Dr. Karen Swidder Lyons, Naval Research Laboratory, for categorizing the retrieved records for query development purposes, and Dr. Richard Carlin, Office of Naval Research, for sponsoring this effort***.)**

KEYWORDS: Text Mining; Bibliometrics; Computational Linguistics; Clustering; Multi-Link; Hierarchical; Information Retrieval; Database Tomography; Electrochemical; Fuel Cells; Batteries; Capacitors; Energy Production; Power Production; Energy Conversion; Energy Storage; Citation Analysis; Scientometrics; Military Requirements

REPORT DOCUMENTATION PAGE

Form Approved OMB No. 0704-0188

1. REPORT DATE (DD-MM-YYYY)	2. REPORT TYPE	3. DATES COVERED (FROM - TO)
14-07-2003	Technical	xx-xx-2000 to xx-xx-2003

4. TITLE AND SUBTITLE
SCIENCE AND TECHNOLOGY TEXT MINING
ELECTROCHEMICAL POWER
Unclassified

5a. CONTRACT NUMBER
5b. GRANT NUMBER
5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Kostoff, Ronald N ;
Tshiteya, Rene ;
Pfeil, Kirstin M ;
Humenik, James A ;

5d. PROJECT NUMBER
5e. TASK NUMBER
5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME AND ADDRESS
Office of Naval Research
800 N. Quincy St.
Arlington, VA22217

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME AND ADDRESS
Office of Naval Research
800 N. Quincy St.
Arlington, VA22217

10. SPONSOR/MONITOR'S ACRONYM(S)
ONR

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
APUBLIC RELEASE
,

13. SUPPLEMENTARY NOTES

14. ABSTRACT
Database Tomography (DT) is a textual database analysis system consisting of two major components: 1) algorithms for extracting multi-word phrase frequencies and phrase proximities (physical closeness of the multi-word technical phrases) from any type of large textual database, to augment 2) interpretative capabilities of the expert human analyst. DT was used to derive technical intelligence from an Electrochemical Power database derived from the Science Citation Index (SCI). Phrase frequency analysis by the technical domain experts provided the pervasive technical themes of the Electrochemical Power database, and the phrase proximity analysis provided the relationships among the pervasive technical themes. Bibliometric analysis of the Electrochemical Power literature supplemented the DT results with author/ journal/ institution publication and citation data. Military requirements for energy and power were overviewed, especially for electrochemical sources and converters.

15. SUBJECT TERMS
Text Mining; Bibliometrics; Computational Linguistics; Clustering; Multi-Link; Hierarchical; Information Retrieval; Database Tomography; Electrochemical; Fuel Cells; Batteries; Capacitors; Energy Production; Power Production; Energy Conversion; Energy Storage; Citation Analysis; Scientometrics; Military Requirements

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19. NAME OF RESPONSIBLE PERSON
			Same as Report (SAR)	86	Kostoff, Ronald kostofr@onr.navy.mil
a. REPORT Unclassified	b. ABSTRACT Unclassified	c. THIS PAGE Unclassified			19b. TELEPHONE NUMBER International Area Code Area Code Telephone Number 703696-4198 DSN -

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std Z39.18

TABLE OF CONTENTS

ABSTRACT

Database Tomography (DT) is a textual database analysis system consisting of two major components: 1) algorithms for extracting multi-word phrase frequencies and phrase proximities (physical closeness of the multi-word technical phrases) from any type of large textual database, to augment 2) interpretative capabilities of the expert human analyst. DT was used to derive technical intelligence from an Electrochemical Power database derived from the Science Citation Index (SCI). Phrase frequency analysis by the technical domain experts provided the pervasive technical themes of the Electrochemical Power database, and the phrase proximity analysis provided the relationships among the pervasive technical themes. Bibliometric analysis of the Electrochemical Power literature supplemented the DT results with author/ journal/ institution publication and citation data. Military requirements for energy and power were overviewed, especially for electrochemical sources and converters.

EXECUTIVE SUMMARY

ES-1. Overview and Military Requirements

A text mining analysis of Electrochemical Power Sources, Converters, and Storage was performed. The technical scope included the generation and conversion of power, and the storage of energy, using electrochemical processes. Military requirements for power sources in general, and Electrochemical power sources in particular, were outlined.

An approximately 175 term query was developed for accessing records from the source SCI database. This query can be used periodically to update the currency of records retrieved.

The bibliometrics sections contained numerical indicators for each bibliometric examined, clustering results for most prolific authors and countries only, and highest frequency elements for each bibliometric. Four technical discipline taxonomies were generated: Keyword non-statistical, Abstract non-statistical, Abstract factor matrix statistical, Abstract multi-link clustering statistical.

Many military applications/ requirements for Electrochemical power sources were identified. Batteries can be used as components of the many military applications listed. They tend to support guidance and control, communications, propulsion, surveillance and detection, fusing, arming, and backup power. Military research is focused on more efficient, smaller, lighter, safer, cheaper, higher power and energy, more reliable, higher longevity, and more safely disposable, batteries.

Fuel cells have the same generic development targets and can potentially be used in many of the same applications as batteries, but they are not as far along in development or implementation. Fuel cells have the potential to be attractive battery

replacements, because their energy storage capability is significantly greater than batteries. Very high power fuel cells are being developed for ship propulsion and ship service power; high power fuel cells are being developed for base stationary power; moderate power fuel cells are being developed for mobile electric power, auxiliary power units, and robotic vehicles; and low power fuel cells are being developed for soldier systems (radios, cooling, heating, weapon systems), battery charging, small robotic vehicles, and remote power. These low power fuel cells have the potential to extend soldier mission times by hours, or possibly days.

Super- or ultra-capacitors are niche storage components. They have higher energy densities than conventional dielectric capacitors, but lower energy densities than batteries or fuel cells. They have higher power densities than fuel cells or batteries, but lower power densities than conventional dielectric capacitors. They are viewed as potentially competitive candidates for modern digital communication devices, which are pulsed and time shared, and involve packet transmission techniques. In their optimal operational frequency range, they can smooth the loads on batteries, thereby increasing capacity and decreasing battery costs and hazards. Their potential ruggedness and reliability are important features.

ES-2. Bibliometrics

ES-2A. Prolific Authors

There were 6985 papers retrieved, 11051 different authors, and 25465 author listings. Of the twenty most prolific authors, seven are from Japan. In fact, thirteen are from the Far East, four are from Europe (Western), two are from North America, and one is from the Middle East. Eighteen are from universities, and two are from research institutes.

Clusters of related authors were generated by two methods: factor matrix and multi-link aggregation. The 253 highest frequency authors were included. Co-author publishing groups were clearly identified, and fine-structure relationships within each group were clearly delineated. Based on name ethnicity alone, intra-country clustering appears to be the dominant form of grouping.

ES-2B. Prolific Journals

The majority of the journals are electrochemistry, with the remainder divided between chemistry and materials. There appear to be three primary groups at the top layer. The Journal of Power Sources contains the most articles by far, and constitutes the first group. The next group consists of the Journal of the Electrochemical Society (JES) and Solid State Ionics (SSI). The next five journals listed (Electrochim Acta, J. Alloy. Compd., Denki Kagaku, J. Appl. Electrochem., J. Electroanal. Chem.) constitute the third group.

ES-2C. Prolific Institutions

Of the twenty most prolific institutions, ten are from Asia, five are from Western Europe, four from the USA, and one from Eastern Europe. Twelve are universities, and the remaining institutions are research institutes.

ES-2D. Prolific Countries

There appear to be three dominant groups in the twenty most prolific countries. The US and Japan constitute the most dominant group, and were the only two countries to have published more than 1000 papers on power-related electrochemistry during the past 8 years. France and China constitute the next group, but had fewer papers combined than either member of the first group. The next seven countries constitute the third group.

Interestingly, unlike all previous DT studies, the United States (US) was not the most prolific country. Japan had more published papers (nearly 18% more) than the US. Overall, Eastern Asia (Japan, China, South Korea, Taiwan), Northern North America (US, Canada), and Western Europe (France, Germany, UK) accounted for most of the Electrochemical Power sources research activity.

ES-2E. Most Cited Authors

Of the twenty most cited authors, five are from Japan, five from the USA, five from Europe (Western), two from Canada, two from Israel, and one from Africa. This is a far different distribution from the most prolific authors, where thirteen were from the Far East. The lists of twenty most prolific authors and twenty most highly cited authors only had four names in common (AURBACH, TARASCON, DAHN, WATANABE). This phenomenon of minimal intersection has been observed in all other text mining studies performed by the first author.

Thirteen of the most cited authors' institutions are universities, four are government-sponsored research laboratories, and three are private companies. The appearance of the companies on this list is another differentiator from the list of most prolific authors.

ES-2F. Most Cited Papers

The Journal of the Electrochemical Society contains the majority of most highly cited papers by far, twelve out of the twenty listed. Most of the journals in which the most highly cited papers are published are fundamental science journals, and most of the topics have a fundamental science theme. Most of the twenty highly cited papers are from the 1990s, with four being from the 1980s, and one extremely highly cited paper being from 1976. This reflects a dynamic research field, with seminal works being performed in the recent past.

Sixteen of the twenty most highly cited papers address issues related to lithium secondary batteries, with the dominant issue theme being lithium insertion/ intercalation

to avoid free-metal formation. Two of the papers address issues related to ceramic fuel cells, with the dominant issue theme being solid oxides for high ionic conductivity. One paper addresses issues related to nickel metal hydride rechargeable batteries.

Thus, the major intellectual emphasis of cutting edge electrochemical power sources research, as evidenced by the most cited papers, is well aligned with the intellectual heritage and performance emphasis, as will be evidenced by the clustering approaches.

ES-2G. Most Cited Journals

The Journal of the Electrochemical Society received as many citations as the next three journals combined. Most of the highly cited journals are electrochemistry, some are materials, some chemistry, with one physics journal represented. Based on all the citation results, there is little evidence that disciplines outside the tightly knit electrochemistry-materials groups relevant to the specific applications are being accessed.

ES-3. Technical Discipline Taxonomies

ES-3A. Keywords Manual Taxonomy

The Keyword non-statistical taxonomy could be divided generically into system-related categories and tech base-related categories.

The main system categories were Sources-Converters, Storage, and Other Applications. Sources-Converters included Fuel Cells, Storage included Batteries and Capacitors (to a much smaller extent), and Other Applications covered many uses of Electrochemical Cells. The main tech-base categories were Electrolytes, Electrodes, Materials, Processes, Properties, Phenomena, Parameters, Experiments, Theory, Macrostructure, Microstructure, and Region.

Fuel cell concepts emphasized were solid oxide, molten carbonate, and polymer electrolyte. While most efforts addressed pure hydrogen input, or reformed hydrocarbon input to supply hydrogen (e.g., the CO-tolerant internal reforming molten carbonate fuel cell), some fuel cell research efforts focused on direct hydrocarbon conversion. These concepts included e.g. solid metal electrolyte and polymer electrolyte fuel cells.

Batteries focused strongly on lithium-ion rechargeable, with substantial effort on nickel-metal hydride and lead-acid, and somewhat less emphasis on rechargeable air-electrode batteries. The lithium-ion battery emphasis was two-fold: increasing electrode intercalation to minimize lithium free-metal formation with non-aqueous electrolytes, and modifying the electrolyte (and electrodes) to increase conductivity and allow higher specific energies. Research emphasis to improve specific energy appeared to be on solid-polymer electrolytes, backed by highly conducting current collectors. The nickel-metal hydride battery is a potential replacement for the balanced performance (specific

energy, specific power, cycle-life, reliability) but environmentally contaminating Ni-Cd battery. Ni-M-H emphasized developing metal hydrides for negative electrodes that can be decomposed and reformed reversibly, and have high hydrogen retention to reduce battery self-discharge rate. The lead-acid battery emphasis appears to be materials, packaging, and fabrication. Much of the associated modeling and simulation reflected in the Keywords was related to lead-acid battery system optimization, focusing on maximizing energy at optimum power density, minimizing internal resistance, retaining maximum charge, and maximizing mechanical strength and cycle life. Environmental impact of lead-acid batteries, mainly disposal and material recyclability, was a research area as well. The rechargeable air-electrode batteries, such as zinc-air, focused on methods to increase charging rates and reduce electrolyte carbonation from air-based carbon compounds.

There was very little effort reflected in capacitors. Relatively high specific power/ specific energy super-capacitors appear to be a key research area, with potential for filling the gap between the high power density low energy density conventional capacitors and the high energy density low power batteries or very high energy density low power density fuel cells. Research focus appears aimed at electrode geometries (larger surface areas) and materials that will allow higher capacitance, and electrolyte materials (and geometries) that will allow the relatively low voltage ceilings (due to potential breakdown across the small separator gap) to be raised.

ES-3B. Abstracts Manual Taxonomy

The Abstract non-statistical taxonomy was generated by visual inspection and manual grouping. The taxonomy was divided into three hierarchical levels, and the global level of emphasis (GLE) of each category and sub-category within the taxonomy was obtained by summing the frequency of the phrases assigned to each category. In the following discussion, the categories in each level are identified, and the sum of the phrase frequencies for each category are shown in parenthesis after the category name.

ES-3B1. Highest Level Taxonomy

The highest taxonomy level consisted of two categories: Electrochemical Converters (17,227) consisted of fuel cell technologies, and Electrochemical Source and Storage Devices (24,804) consisted of battery and electrochemical capacitor technologies.

An evaluation of the <u>current status</u> of these technologies showed that electrochemical source/storage devices provided power to a wider variety of applications (especially for small / portable electronic devices) compared to fuel cells. As the numbers of applications requiring portable power continue to increase rapidly, near term solutions are required, and batteries currently seem to offer the most feasible solution. Hence, the higher level of emphasis in that category.

The literature indicated that Fuel Cell research was substantial. The potential of fuel cell technology is promising, but many technical and economic issues such as the need for expensive catalysts (in low temperature fuel cells), the large size / weight / miniaturization, and the corrosion / breakdown of components (for high temperature Fuel Cells) would need to be resolved for it to become more competitive. Another issue is the lack of infrastructure for some of the Fuel Cells' fuels (methanol, ethanol, and hydrogen).

Supercapacitors remain ideal for delivering high power over a short duration, but are most useful when combined with other electric power sources (e.g., to provide extra power boost in an electric application). Technical advances were still required before higher energy density systems could be mass-produced.

ES-3B2. Second Level Taxonomy

ES-3B2a. Fuel Cells

Fuel Cell research addressed the following tech base areas: system components and component configurations (4,038); properties and characteristics (2,683); sources / fuels (2,385); materials (2,358); conversion processes (1,363); conversion byproducts (1,011); operating conditions (885); and potential applications (267).

In general, Fuel Cell research was aimed at improving the performance of Fuel Cells (for a wider variety of applications) while lowering the cost (manufacturing, operation, and maintenance) of the systems. Advances in the individual Fuel Cell components could help achieve many of those objectives. For example:

- Use of porous electrodes to increase electrode surface area, and subsequently, improve electrode current density and overall fuel cell performance;
- Use of polymer electrolytes to reduce costly corrosion and electrolyte loss problems, and to improve hydrogen ion conductivity (which improves fuel cell performance);
- Development of electrocatalysts to improve the rate of electrode reactions / exchange current density.
-

The importance of advances in component technologies was reflected by the high GLE for that category.

The Fuel Cell properties / characteristics (such as ionic conductivity, current density, corrosion rate), sources / fuels (such as hydrogen, methane, natural gas, hydrocarbons, methanol, ethanol), and materials (such as nafion, zironia, platinum, ceramics) seemed to have equal amounts of research according to the GLE. All three of these categories provide critical supporting / enabling technologies for improving Fuel Cells from a performance or cost perspective.

Other important Fuel Cell categories were conversion processes, byproducts, and operating conditions. Lastly, there was little emphasis on applications. Current Fuel

Cell technologies did not address a high variety of applications (based on range of required energy or rate of power output).

Most of the recent research on Fuel Cell technologies focused on solid oxide and molten carbonate Fuel Cells, due to their high operating temperatures and consequently higher efficiencies and relaxed reforming requirements Polymer electrolyte / Proton Exchange Membrane (PEM) Fuel Cells were also being pursued, but at a lower level. The Fuel Cells' main technical areas of interest were the same as described in the Keyword taxonomy. For solid oxide Fuel Cells (SOFC) electrolyte doping, synthesis and characterization of various anode / cathode materials, and the development of low temperature SOFC seemed to be the focal points of research. Molten carbonate fuel cell (MCFC) research was focused on the development of novel cathode materials (often a nickel based alloy for corrosion resistance between the cathode and current collector plate), and the selection of anode catalyst for the reforming reaction. Polymer electrolyte and Proton Exchange Membrane (PEM) Fuel Cell research was developing new techniques for preparing the catalyst layer in the electrolyte, developing new fabrication methods for composite membranes, and was assessing the physical and morphological characteristics and electrochemical behavior of various catalysts (i.e. PtRu/C).

ES-3B2b. Batteries

Batteries research addressed the following tech base areas: materials (7,850), properties and characteristics (4,643), component technologies (4,531), processes and phenomena (2,658), types (2,195), and applications (1,121)

Materials seemed to be an integral part of advancing battery technologies. The enhanced properties and characteristics of novel materials, and their application in the electrodes, electrolytes, or at the interface were projected to result in lighter, higher energy capacity, less expensive batteries. For example, research was being conducted on:

- porous metals / high surface area materials for NiCd / NiMH electrode;
- hydrogen absorbing metals for NiMH anodes;
- alloy powders for higher energy capacity and improved cycle life;
- carbon coated silicon for anodes to improved cycleability;
- analyses of reactions at electrode-electrolyte (non-aqueous) interface;
- polymer electrolytes;
- carbon anodes / composite anodes;
- microwave synthesis of electrode materials;

Materials were expected to improve properties and characteristics of batteries and battery components (electrodes and electrolytes) with similar impacts. This was reflected by the equal amounts of emphasis in each category, according to the GLE numbers.

Little was discussed in the literature about the basic theory behind the electrochemical processes on which batteries are based, other than material synthesis and characterization. There were some efforts in the areas of modeling and simulation, mainly focused on lead-acid batteries. There was a modest amount of the literature that focused on applications of batteries.
Technical area research for battery technologies had a strong focus on secondary lithium batteries (Lithium Ion, Lithium Polymer, etc.). Research included the investigation of carbon-carbon (C-C) composite as an anode material for lithium ion batteries. Lithium-ion cells made with the C-C composite anode showed many advantages, such as excellent performance and enhanced safety. Other research focused on improving the charge-discharge characteristics of polyaniline films used as cathodes in lithium batteries and improving the ionic conductivity of crystalline polymer electrolytes. Lead-acid batteries also continued to be a major thrust area. Research there focused on battery additives and their influence on the separator behavior. There was research on the use of gas-recombining noble metal catalysts in valve regulated lead acid batteries. Also, researchers were developing high rate discharge (HRD) lead-acid batteries, geared for automobile applications. Nickel Metal Hydride (NiMH) batteries constituted another major thrust area. Research was conducted to develop new / improved methods for synthesizing and characterizing materials. For example, surface modification of hydrogen storage alloy electrodes by the hot-charging treatment was being investigated. New hydrogen storage materials were being developed. Gelatin-pretreated graphite anodes were being tested to see if they could improve the irreversible loss of charge, reversible capacity and efficiency of NiMH batteries.

ES-3B2c. Electrochemical Capacitors

Electrochemical capacitors research addressed the following tech base areas: Properties and characteristics (604); component technologies (568); materials (435); and types (199). The major concerns were associated with improving the energy density and power density, hence the higher GLE for that category. Potential solutions for achieving this included improving components (dielectric / gel electrolyte, solid electrolyte; polarizable electrodes/ composite electrodes) and improvements based on component materials (glassy carbon, carbon fibers, aerogels, thin films).

A focal point of electrochemical capacitor research was the chemical synthesis and characterization of various materials (i.e. MoxSy (CO)(n) and Mo-x(CO)(n)) and their possible application as catalysts for oxygen reduction reaction. Other electrochemical capacitor thrust areas included: the use of industrial carbon blacks (CBs) materials in electrochemical supercapacitors because of their high specific areas; the development of thin film supercapacitors using a sputtered RuO2 electrode; the development of carbon nanotubes/RuO2 electrodes for electrochemical capacitors; the synthesis, structure-property characterization, and performance of carbon aerogels; and the fabrication and application of Cu-carbon composite (prepared from sawdust) to electrochemical capacitor electrodes.

The absence of any categories/ sub-categories in this taxonomy should not be interpreted that S&T efforts are not being pursued in those areas. The correct interpretation is that within the frequency threshold constraints of the electrochemical database, mid-high frequency phrases related to these categories do not appear.

Overall, the Abstract manual taxonomy and GLE suggest that there is, and continues to be, a larger amount of research activity related to batteries than fuel cells, and electrochemical capacitor research activity is extremely small (relative to both batteries and fuel cells).

ES-3C. Abstract Factor Matrix Taxonomy

The Abstract factor matrix was based on the eigenvalues having a floor of unity, and the matrix had twenty factors. These factors were analyzed in detail for technical themes, and were further aggregated to form a taxonomy. The Abstract factor matrix statistical taxonomy could be represented by a three level hierarchy. The highest level taxonomy (level 1) consisted of Fuel Cells, Batteries, and Capacitors. The next highest level taxonomy (level 2), and its level 3 components, is:

Fuel Cells
*SOFC – improve electric conductivity and thin film properties, and reduce thermal expansion, to increase current density and maximum power density.
*DMFC – catalysts and hydrogen storage alloys to improve hydrocarbon oxidation and hydrogen storage.
*Materials – materials for electrochemical systems, including fuel cells.
*Diagnostics – diagnostic techniques used to characterize phenomena and properties of electrochemical systems, including fuel cells.

Batteries
*Lithium-ion – rechargeable.
-Intercalation
-Thin films
-Conductivity
-Reversibility
-Fabrication
*Lead-Acid – modeling. simulation, and performance characteristics measurement.
*Ni-M-H – hydrogen storage alloys structure and reactions.
*Materials – materials for electrochemical systems, including batteries.
*Diagnostics – diagnostic techniques used to characterize phenomena and properties of electrochemical systems, including batteries.

Capacitors
*Thin Films – capacitance and energy density of thin-film capacitors.

ES-3D. Abstract Multi-Link Clustering Taxonomy

The Abstract multi-link clustering tree structure was divided into fourteen component branches. Each branch was analyzed in detail for thematic emphasis, and further aggregated to form a taxonomy. In the Abstract multi-link clustering statistical taxonomy, the main research focal points appear to be solid oxide fuel cells (increasing electric/ ionic conductivity, direct hydrocarbon oxidation) and lithium secondary batteries (improving electrode intercalation, increasing electrolyte conductivity, increasing discharge capacity/ reversibility). Research also emphasizes other direct hydrocarbon fuel cells and molten carbonate, and other secondary batteries, such as Ni-M-H and lead-acid.

ES-4. Final Observations

The different statistical and non-statistical taxonomies generated above used different methodologies and some different phrases. Therefore, the results are not directly comparable. A taxonomy that reflects the levels of effort and specific research thrusts would have the structure of the non-statistical Abstract field taxonomy. It would reflect high emphasis on the solid oxide fuel cells and rechargeable lithium batteries, especially the sub-thrusts identified in the statistical clustering approaches. It would also reflect emphasis on other rechargeable battery approaches, such as Ni-M-H, and on direct hydrocarbon fuel cells. Its capacitor component would be low, with emphasis on thin-film voltage-breakdown-resistant supercapacitors.

Two final observations on the technical results. Based on reading a large number of the retrieved SCI Abstracts, and examining the phrase listings and taxonomy results, there appears to be a large imbalance between theory and experiment. The discipline appears almost Edisonian in nature. In addition, there is little evidence of extrapolation of concepts and insights from other technical disciplines. The research appears very parochial. Some of the hybrid literature-based discovery/ multi-discipline workshop techniques (1), performed at the initiation of research projects, could systematically access this extra-disciplinary information.

This paper has presented a number of advantages of using DT and bibliometrics for deriving technical intelligence from the published literature. Large amounts of data can be accessed and analyzed, well beyond what a finite group of expert panels could analyze in a reasonable time period. Preconceived biases tend to be minimized in generating roadmaps. Compared to standard co-word analysis, DT uses full text, not index words, and can make maximum use of the rich semantic relationships among the words. It also has the potential of identifying low occurrence frequency but highly theme related phrases that are 'needles-in-a-haystack', a capability unavailable to any of the other co-occurrence methods.

Combined with bibliometric analyses, DT identifies not only the technical themes and their relationships, but relationships among technical themes and authors, journals, institutions, and countries. Unlike other roadmap development processes, DT generates

the roadmap in a 'bottom-up' approach. Unlike other taxonomy development processes, DT can generate many different types of taxonomies (because it uses full text, not key words) in a 'bottom-up' process, not the typical arbitrary 'top-down' taxonomy specification process. Compared to co-citation analysis, DT can use any type of text, not only published literature, and it is a more direct approach to identifying themes and their relationships.

The maximum potential of the DT and bibliometrics combination can be achieved when these two approaches are combined with expert analysis of selected portions of the database. If a manager, for example, wants to identify high quality research thrusts as well as science and technology gaps in specific technical areas, then an initial DT and bibliometrics analysis will provide a contextual view of work in the larger technical area; i.e., a strategic roadmap. With this strategic map in hand, the manager can then commission detailed analysis of selected abstracts to assess the quality of work done as well as identify work that needs to be done (promising opportunities).

1. INTRODUCTION

Science and technology are assuming an increasingly important role in the conduct and structure of domestic and foreign business and government. In the highly competitive civilian and military worlds, there has been a commensurate increase in the need for scientific and technical intelligence to insure that one's perceived adversaries do not gain an overwhelming advantage in the use of science and technology. While there is no substitute for direct human intelligence gathering, there have become available many techniques that can support and complement it. In particular, techniques that identify, select, gather, cull, and interpret large amounts of technological information semi-automatically can expand greatly the capabilities of human beings in performing technical intelligence.

One such technique is DT (2,3, 4), a system for analyzing large amounts of textual computerized material. It includes algorithms for extracting multi-word phrase frequencies and phrase proximities from the textual databases, coupled with the topical expert human analyst to interpret the results and convert large volumes of disorganized data to ordered information. Phrase frequency analysis (occurrence frequency of multi-word technical phrases) provides the pervasive technical themes of a database, and the phrase proximity (physical closeness of the multi-word technical phrases) analysis provides the relationships among pervasive technical themes, as well as among technical themes and authors/journals/institutions/countries, etc. The present paper describes use of the DT process, supplemented by literature bibliometric analyses, to derive technical intelligence from the published literature of Electrochemical Power science and technology.

Electrochemical Power, as defined by the authors for this study, is the generation and conversion of power, and the storage of energy, using electrochemical processes. Since one of the key outputs of the present study is a query that can be used by the community to access relevant Electrochemical Power documents, a recommended query based on this study is presented in total. This query serves as the operational definition of Electrochemical Power, and its development is discussed in detail in the database generation section.

ELECTROCHEMICAL POWER QUERY

(fuel cell* or sofc* or pemfc* or dmfc* or ultracapacitor* or supercapacitor* or pseudocapacitor* or (capacitor* same (electrochemical or electrolyte* or double-layer)) or ((battery or batteries) same (lithium or li or electrode* or anode* or cathode* or capacity or material* or electrochemical or charge or charging or discharge* or discharging or rechargeable or electrolyte* or lithium or li or lithium-ion or nickel or metal hydride* or lead-acid or alloy*)) or ((lithium or li) same (electrochemical or discharge* or discharging or electrode* or liclo4 or rechargeable or cycling or reversible or insertion or mah or intercalation)) or (electrochemical same (discharge* or discharging or hydrogen storage or mah)) or (hydrogen storage same (alloy* or electrode*)) or (limn2o4 same electrode*) or (lipf6 same electrolyte*) or (charge-discharge same electrode*) or ((discharge capacity or metal hydride*) same electrode*) or (electrolyte* same lsgm) or

(hydrogen same storage alloy*) or (nafion same polymer*) or (ptru same co) or (ruo2 same electrode*)) NOT(((electrode* or hydrogen or discharge*) same plasma*) or (discharge* same gas) or dna or assay* or biosensor* or rats or blood or capillary or protein* or mercury or clinical or amino or hydrogen peroxide or paste or corona or tissue* or helium or ascorbic acid or receptor* or chromium or radiation or bacteria* or plant* or extracellular or antenna* or magnetron or drug* or vivo or hydrolysis or ml or amperometric or care or cd or buffer or silicon or stress or sensor* or rf or filter* or switching or detection limit* or inhibition* or ar or ms or electrostatic or phi or monolayer* or gate* or sheath* or gc or depletion or combustion or serum* or toxicity or converter* or chromatography or radical* or oil* or generator* or target* or gap* or excitation* or environmental or glow* or ring or rings or diet* or pretreatment* or space charge* or amine* or ultrasound or lamp* or scan rate* or health* or solar or fe2 or reflection* or electromagnetic or carboxylic or deep or diode* or synthetic* or acetic acid or collision* or moiety or dimeric or titanate* or carbon steel* or curvature* or lithium chloride or coercive field or network* or hydrodynamic* or tris or mutant* or backbone* or decay* or monomer* or outcome* or driving or contamination or spatial or cmos or mediator* or excited or led or self-assembled or nitric oxide or i-v or array* or mmol or dt or waste* or aromatic or epitaxial or atomic force microscopy or differential pulse or viscosity or sorption or pk or native or shifts or recording* or adhesion* or dye* or surfactants)

To execute the study reported in this paper, a database of relevant Electrochemical Power articles is generated using the iterative search approach of Simulated Nucleation (5, 6). Then, the database is analyzed to produce the following characteristics and key features of the Electrochemical Power field: recent prolific Electrochemical Power authors; journals that contain numerous Electrochemical Power papers; institutions that produce numerous Electrochemical Power papers; keywords most frequently specified by the Electrochemical Power authors; authors, papers and journals cited most frequently; pervasive technical themes of Electrochemical Power; and relationships among the pervasive themes and sub-themes.

What is the importance of applying DT and bibliometrics to a topical field such as Electrochemical Power? The roadmap, or guide, of this field produced by DT and bibliometrics provides the demographics and a macroscopic view of the total field in the global context of allied fields. This allows specific starting points to be chosen rationally for more detailed investigations into a specific topic of interest. DT and bibliometrics do not obviate the need for detailed investigation of the literature or interactions with the main performers of a given topical area in order to make a substantial contribution to the understanding or the advancement of this topical area, but allow these detailed efforts to be executed more efficiently. DT and bibliometrics are quantity-based measures (number of papers published, frequency of technical phrases, etc.), and correlations with intrinsic quality are less direct. The direct quality components of detailed literature investigation and interaction with performers, combined with the DT and bibliometrics analysis, can result in a product highly relevant to the user community.

2. BACKGROUND

2.1 Military Requirements for Energy and Power

Fundamental to the operation of all advanced modern militaries is availability of energy and power supplies that will remove roadblocks to successful conduct of strategic and tactical missions. Different missions require far different power supplies, with different operating characteristics.

To compare the diversity of available and potential power supplies with the myriad military missions and operations possible, some type of taxonomic scheme is required. One categorization revolves around whether humans are located in proximity of the power supply during the mission. Another is by geospatial location (space, atmosphere, land, sea, sub-surface) of the power supply during the mission. A third categorization is by the technology that uses the power supply (e.g., propulsion, communications, heating). A fourth categorization is by the type of fuel source (e.g., fossil, solar, nuclear, wind, etc). A fifth type of categorization is by the type of converter (e.g., heat cycle, direct conversion). Because of space limitations, this section will concentrate on the first two taxonomies.

The first taxonomy is power supplies in remote missions (where humans are not involved in-situ) and in direct missions (where humans are involved in-situ). Remote operations (e.g., space, underwater, underground, and land/ air-based robotic systems) can be further sub-divided into short-term (typically weapons launches) and long-term (typically surveillance, communications nodes). Long-term remote missions need supplies that are highly reliable (no maintenance required), long-lived, and retain performance over many cycles. While cost and efficiency are important, especially where numerous detectors with large data outputs are required, cost and efficiency could be traded off for reliability, and absence of moving parts is usually considered a positive factor. Safety issues, such as environmental hazards, are less important for remote operations than where humans are involved in-situ. Long-term space missions require supplies that are lightweight (because of launch costs), launch survivable, low-G compliant, and survivable in the unique space environment (high radiation bands, large temperature swings, potential low pressure operation). Long-term buried or covert supplies (e.g., for detectors) do not have the critical weight limitation of space systems, but could be subject to harsh environmental conditions (e.g., corrosion-generating), and could have more stringent reduced signature requirements (thermal, acoustic, magnetic). Short-term remote applications (e.g., smart munitions) might have long shelf life requirements, high stress operation requirements (e.g., high-G, high temperature swings, high pressure, high vibration, high shock, high radiation, high magnetic fields), and high power density requirements, but long cycle repetition requirements would be reduced substantially. For direct operations, safety and hazard reduction considerations increase substantially, and high stress environments decrease, sometimes drastically.

The second categorization of missions discussed is geo-spatial. For space missions, power is used for vehicle and weapons propulsion, pulsed weapons, communications, surveillance, and housekeeping. Vehicle and weapons propulsion tend to be moderate/ short term high power density, pulsed weapons tend to be very high power very short term, and communications and surveillance are relatively low power and long term (with operating cycles that can range from short to long term). Other criteria for space operations were presented above.

For atmospheric missions, power is used for many of the same generic applications as space, with the major additions of combat and transport of people and materiel. Missions can be remote or direct. For both atmospheric and space missions, weight and size assume more importance than for terrestrial missions, with the exception of man-portable systems.

For stationary land-based direct missions, power is used for base maintenance operations (heating, cooling, lights, appliances, etc), communications, surveillance, local vehicle propulsion, and supply. For stationary land-based remote missions, power is used mainly for surveillance and communications, and for propulsion of robotic systems. For mobile land-based direct missions, power is used for propulsion, communications, and surveillance. For the specific case of the individual land-based warrior, power is generically required for the computer/ radio subsystem, the software subsystem, the integrated helmet assembly subsystem, and the weapon subsystem. For mobile land-based remote missions, power is used for weapons propulsion, guidance, surveillance, and communications. In the above, power production on-board a flying weapon is considered mobile remote.

For sea surface and undersea applications, the types of power requirements are comparable to those for a combination of air and land-based systems (e.g., combat, troop and materiel transport, short pulsed high power weapons, moderate pulse weapons), but the operating environment tends to be somewhat harsher (e.g., especially saline corrosion). In addition, long-term manned undersea missions tend to have higher reliability requirements more approximating those of space missions, while at the same time experiencing the constraints required for direct missions.

In general, evolving military applications require decreases in size and weight, especially for space, aircraft, and individual soldier or small team applications. For large volumes of power supply applications, such as munitions and radios, reduced cost becomes an important factor. For either weight or size reduction, or increased mission longevity, increase in energy and power density becomes important. Where people are involved, increased safety is important, and for long-term operations, environmental compliance is important. High reliability is of importance, especially where maintenance is not possible during the course of the mission (space, weapons flight, covert surveillance). Where maintenance is possible, ease of maintenance and supportability are important power supply considerations. In some militaries, limitations are placed on the types of fuels that can be used (e.g., diesel, JP-type fuels). The trend is also toward faster vehicles and weapons. Aerodynamics dictates power requirements will increase

nonlinearly with speed, and for fixed size vehicles, larger power supplies will be required.

2.2 Characteristics of Electrochemical Energy and Power

There are three main electrochemical source/ converter/ storage systems: batteries, fuel cells, and capacitors. Relative to heat engines, they have far fewer moving parts, eliminate the need for a thermal conversion step, and tend to be more reliable with lower acoustic and thermal signatures. Relative to renewable sources, they have higher energy and power densities (excluding fission or fusion as renewable sources).

2.3 Electrochemical Energy and Power for Military Applications

Batteries can be used as components of the many military applications listed above. They tend to support guidance and control, communications, propulsion, surveillance and detection, fusing, arming, and backup power. Military research is focused on more efficient, smaller, lighter, safer, cheaper, higher power and energy, more reliable, higher longevity, and more safely disposable, batteries.

Fuel cells have the same generic development targets and can potentially be used in many of the same applications as batteries, but they are not as far along in development or implementation. Fuel cells have the potential to be attractive battery replacements, because their energy storage capability is significantly greater than batteries. Very high power fuel cells are being developed for ship propulsion and ship service power; high power fuel cells are being developed for base stationary power; moderate power fuel cells are being developed for mobile electric power, auxiliary power units, and robotic vehicles; and low power fuel cells are being developed for soldier systems (radios, cooling, heating, weapon systems), battery charging, small robotic vehicles, and remote power. These low power fuel cells have the potential to extend soldier mission times by hours, or possibly days.

Super- or ultra-capacitors are niche storage components. They have higher energy densities than conventional dielectric capacitors, but lower energy densities than batteries or fuel cells. They have higher power densities than fuel cells or batteries, but lower power densities than conventional dielectric capacitors. They are viewed as potentially competitive candidates for modern digital communication devices, which are pulsed and time shared, and involve packet transmission techniques. In their optimal operational frequency range, they can smooth the loads on batteries, thereby increasing capacity and decreasing battery costs and hazards. Their potential ruggedness and reliability are important features.

2.4 Text Mining Overview

The information sciences background for the approach used in this paper is presented in (7). This reference shows the unique features of the computer and co-word-based DT process relative to other roadmap techniques. It describes the two main roadmap

categories (expert-based and computer-based), summarizes the different approaches to computer-based roadmaps (citation and co-occurrence techniques), presents the key features of classical co-word analysis, and shows the evolution of DT from its co-word roots to its present form.

The DT method in its entirety requires generically three distinct steps. The first step is identification of the main themes of the text being analyzed. The second step is determination of the quantitative and qualitative relationships among the main themes and their secondary themes. The final step is tracking the evolution of these themes and their relationships through time. The first two steps are summarized in 2.1.1 and 2.1.2. . Time evolution of themes has not yet been studied.

At this point, a variety of different analyses can be performed. For databases of non-journal technical articles (2), the final results have been identification of the pervasive technical themes of the database, the relationship among these themes, and the relationship of supporting sub-thrust areas (both high and low frequency) to the high-frequency themes. For the more recent studies in which the databases are journal article abstracts and associated bibliometric information (authors, journals, addresses, etc), the final results have also included relationships among the technical themes and authors, journals, institutions, etc (7-11).

2.4.1. First Step

The frequencies of appearance in the total text of all single word phrases (e.g., Matrix), adjacent double word phrases (e.g., Metal Matrix), and adjacent triple word phrases (e.g., Metal Matrix Composites) are computed. The highest frequency significant technical content phrases are selected by topical experts as the pervasive themes of the full database.

2.4.2. Second Step

2.4.2.1. Numerical Boundaries

For each theme phrase, the frequencies of phrases within +/-M (nominally 50) words of the theme phrase are computed for every occurrence of the theme phrase in the full text, and a phrase frequency dictionary is constructed. This dictionary contains the phrases closely related to the theme phrase. Numerical indices are employed to quantify the strength of this relationship. Both quantitative and qualitative analyses are performed by the topical expert for each dictionary (hereafter called cluster) yielding, among many results, those sub-themes closely related to and supportive of the main cluster theme.

Threshold values are assigned to the numerical indices, and these indices are used to filter out the phrases most closely related to the cluster theme. However, because numbers are limited in their ability to portray the conceptual relationships among themes and sub-themes, the qualitative analyses of the extracted data by the topical experts

have been at least as important as the quantitative analyses. The richness and detail of the extracted data in the full text analysis allow an understanding of the theme inter-relationships not heretofore possible with previous text abstraction techniques (using index words, key words, etc.).

2.4.2.2. Semantic Boundaries

The approach is conceptually similar to 2.4.2.1, with the difference being that semantic boundaries are used to define the co-occurrence domain rather than numerical boundaries. The only semantic boundaries used for the present studies were paper Abstract boundaries. Software is being developed that will allow paragraphs or sentences to be used as semantic boundaries.

It is an open question as to whether semantic boundaries or numerical boundaries provide more accurate results. The elemental messages of text are contained in concepts or thoughts. Sentences or paragraphs are the vehicles by which the concepts or thoughts are expressed. The goal of text mining is to usually quantify relationships occurring in the concepts or thoughts, not in the fragments of their vehicles of expression. In particular, while intra-sentence relationships will be very strong, they may be overly restrictive for text mining purposes, and many cross-discipline relationships can be lost by adhering to intra-sentence relationships only. Intra-paragraph relationships are more inclusive and reasonable. For journal paper Abstracts of the type found in SCI, many Abstracts constitute a single paragraph.

2.5 Unique Study Features

The study reported in the present paper is in the latter (journal article abstract) category. It differs from the previous published papers in this category (7-11) in two respects. First, the topical domain (Electrochemical Power) is completely different. Second, a much more rigorous statistically-based technical theme clustering approach is used. Third, bibliometric clustering is presented for two database fields: authors and countries.

3. DATABASE GENERATION

The key step in the Electrochemical Power literature analysis is the generation of the database to be used for processing. There are three key elements to database generation: the overall objectives, the approach selected, and the database used. Each of these elements is described.

3.1 Overall Study Objectives

The main objective was to identify global S&T that had both direct and indirect relations to Electrochemical Power. A sub-objective was to estimate the overall level of global effort in Electrochemical Power S&T, as reflected by the emphases in the published literature.

3.2 Databases and Approach

For the present study, the SCI database was used. The approach used for query development was the DT-based iterative relevance feedback concept (5).

3.2.1 Science Citation Index (12)

The database consists of selected journal records (including authors, titles, journals, author addresses, author keywords, abstract narratives, and references cited for each paper) obtained by searching the web version of the SCI for Electrochemical Power articles. At the time the data was extracted for the present paper (mid-2001), the version of the SCI used accessed about 5600 journals (mainly in physical, engineering, and life sciences basic research).

The SCI database selected represents a fraction of the available Electrochemical Power (mainly research) literature, that in turn represents a fraction of the Electrochemical Power S&T actually performed globally (13). It does not include the large body of classified literature, or company proprietary technology literature. It does not include technical reports or books or patents on Electrochemical Power. It covers a finite slice of time (1991 to mid-2001). The database used represents the bulk of the peer-reviewed high quality Electrochemical Power research, and is a representative sample of all Electrochemical Power research in recent times.

To extract the relevant articles from the SCI, the Title, Keyword, and Abstract fields were searched using Keywords relevant to Electrochemical Power, although different procedures were used to search the Title and Abstract fields (5). The resultant Abstracts were culled to those relevant to Electrochemical Power. The search was performed with the aid of two powerful DT tools (multi-word phrase frequency analysis and phrase proximity analysis) using the process of Simulated Nucleation (5).

An initial query of Electrochemical Power-related terms produced two groups of papers: one group was judged by domain experts to be relevant to the subject matter, the other

was judged to be non-relevant. Gradations of relevancy or non-relevancy were not considered. An initial database of Titles, Keywords, and Abstracts was created for each of the two groups of papers. Phrase frequency and proximity analyses were performed on this textual database for each group. The high frequency single, double, and triple word phrases characteristic of the relevant group, and their boolean combinations, were then added to the query to expand the papers retrieved. Similar phrases characteristic of the non-relevant group were effectively subtracted from the query to contract the papers retrieved. The process was repeated on the new database of Titles, Keywords, and Abstracts obtained from the search. A few more iterations were performed until the number of records retrieved stabilized (convergence). The final phrase-based query used for the Electrochemical Power study was shown in the Introduction.

The authors believe that queries of these magnitudes and complexities are required when necessary to provide a tailored database of relevant records that encompasses the broader aspects of target disciplines. In particular, if it is desired to enhance the transfer of ideas across disparate disciplines, and thereby stimulate the potential for innovation and discovery from complementary literatures (1), then even more complex queries using Simulated Nucleation may be required.

4. RESULTS

The results from the publications bibliometric analyses are presented in section 4.1, followed by the results from the citations bibliometrics analysis in section 4.2. Results from the DT analyses are shown in section 4.3. The SCI bibliometric fields incorporated into the database included, for each paper, the author, journal, institution, and Keywords. In addition, the SCI included references for each paper.

The bibliometrics sections (4.1, 4.2) have three components. Some numerical indicators are presented for each bibliometric examined. Clustering results, that portray cohesive groups, are presented for most prolific authors only. Finally, the highest frequency bibliometrics (e.g., most prolific author, most prolific country) are presented for each bibliometric, and discussed.

The DT sections contain three components. First, the high frequency Keywords are grouped into 'natural' categories, and the picture they provide of the Electrochemical Power literature (research, open literature, unclassified, non-proprietary) is described. Second, the high frequency phrases from the Abstracts are grouped into 'natural' categories, and the picture they provide of the Electrochemical Power literature is presented. Third, the high numerical indicator phrases from the proximity analyses of the Abstracts and other portions of the database (author names, article titles, journal names, author addresses) are grouped into 'natural' categories, and the picture they provide of the Electrochemical Power literature is shown. The meaning of the term 'natural' is that these categories were not prescribed beforehand. From observation of the hundreds of different phrases and their frequencies, categories useful for interpreting and describing the main literature findings appeared to emerge.

The analytical approaches taken for the first three components (Keyword phrase frequency, Abstract phrase frequency, Abstract phrase proximity) are based on their fundamental data structures. The Keyword and Abstract phrase frequencies are essentially quantity measures. They lend themselves to 'binning', and addressing adequacies and deficiencies in levels of effort. They do not contain relational information, and therefore offer little insight into S&T linkages.

The phrase proximity results are essentially relational measures, although some of the proximity results imply levels of effort that support specific S&T areas. The phrase proximity results mainly offer insight into S&T linkages, and have the potential to help identify innovative concepts from disparate disciplines (1). Thus, the Keyword and Abstract phrase frequency analyses will be addressed to adequacy of effort, and the phrase proximity analyses will be addressed to relationships primarily and supporting levels of effort secondarily.

4.1 Publication Statistics on Authors, Journals, Organizations, Countries

The first group of metrics presented is counts of papers published by different entities. These metrics can be viewed as output and productivity measures. They are not direct

measures of research quality, although there is some threshold quality level inferred, since these papers are published in the (typically) high caliber journals accessed by the SCI.

4.1.1 Prolific Electrochemical Power Sources Authors

Previous DT/ bibliometrics studies were conducted of the technical fields of: 1) Near-earth space (NES) (8); 2) Hypersonic and supersonic flow over aerodynamic bodies (HSF) (7); 3) Chemistry (JACS) (9) as represented by the Journal of the American Chemical Society; 4) Fullerenes (FUL) (10); 5) Aircraft (AIR) (11); 6) Hydrodynamic flow over surfaces (HYD); 7) Electric power sources (EPS); and 8) the non-technical field of research impact assessment (RIA). Overall parameters of these studies from the SCI database results and the current electrochemical study are shown in Table 1.

TABLE 1 - DT STUDIES OF TOPICAL FIELDS

TOPICAL AREA	**NUMBER OF SCI ARTICLES**	**YEARS COVERED**
1) NEAR-EARTH SPACE (NES)	5480	1993-MID 1996
2) HYPERSONICS (HSF)	1284	1993-MID 1996
3)CHEMISTRY (JACS)	2150	1994
4) FULLERENES (FUL)	10515	1991-MID 1998
5) AIRCRAFT (AIR)	4346	1991-MID 1998
6) HYDRODYNAMICS (HYD)	4608	1991-MID 1998
7) ELECTRIC POWER SOURCES (EPS)	20835	1991-BEG 2000
8) RESEARCH ASSESSMENT (RIA)	2300	1991-BEG 1995
9) ELECTROCHEMICAL POWER SOURCES (ECHEM)	6985	1993 – MID 2001

These studies yielded: 1) 3.37 authors per paper for the NES results; 2) 2.63 authors per paper for the HSF results; 3) 3.79 authors per paper for the JACS results; 4) 3.92 authors per paper for the FUL results; 5) 2.09 authors per paper for the AIR results; 6) 2.29 authors per paper for the HYD results; and 7) 2.90 authors per paper for the EPS results. A previous study on the non-technical field of research impact assessment (RIA) (9) yielded about 1.68 authors per paper.

In the present Electrochemical Power study, two types of computational linguistics analysis were performed on the author data field in the database. First, a frequency count of author appearances was made from the author field in the database, to identify the most prolific authors. Second, a clustering analysis was performed on the list of author appearances, to identify tightly-knit multi-author groups. The clustering methodology (also used in the analysis of the Abstract free text field to generate technology taxonomies) is described in Appendix 1.

4.1.1.1 Author Frequency Results

There were 6985 papers retrieved, 11051 different authors, and 25465 author listings. The occurrence of each author's name on a paper is defined as an author listing. While the average number of listings per author is about 2.3, the twenty most prolific authors (see Table 2) have listings more than an order of magnitude greater than the average. The number of papers listed for each author are those in the database of records extracted from the SCI using the query, not the total number of author papers listed in the source SCI database.

TABLE 2 – MOST PROLIFIC AUTHORS
(present institution listed)

AUTHOR NAME	INSTITUTION	COUNTRY	# PAPERS
DAHN, JR	DALHOUSIE UNIV	CANADA	67
TARASCON, JM	UNIV PICARDIE	FRANCE	53
WANG, QD	ZHEJIANG UNIV	CHINA	51
LEI, YQ	ZHEJIANG UNIV	CHINA	46
LIU, HK	UNIV WOLLONGONG	AUSTRALIA	44
DOU, SX	UNIV WOLLONGONG	AUSTRALIA	44
SCROSATI, B	UNIV ROMA LA SAPIENZA	ITALY	43
LEE, JY	NATIONAL UNIV SINGAPORE	SINGAPORE	42
KUMAGAI, N	IWATE UNIV	JAPAN	41
YAMAMOTO, O	AICHI INST TECHNOLOGY	JAPAN	40
YOSHIO, M	SAGA UNIV	JAPAN	40
AURBACH, D	BAR ILAN UNIV	ISRAEL	38
UCHIDA, I	TOHOKU UNIV	JAPAN	37
WATANABE, M	UNIV YAMANASHI	JAPAN	37
CHEN, LQ	CHINESE ACAD SCIENCE	CHINA	36
TAKEDA, Y	MIE UNIV	JAPAN	36
PASSERINI, S	ENEA	ITALY	35
TIRADO, JL	UNIV CORDOBA	SPAIN	33
IWAKURA, C	UNIV OSAKA PREFECTURE	JAPAN	32
WHITE, RE	UNIV SOUTH CAROLINA	USA	32

Of the twenty most prolific authors listed in Table 2, seven are from Japan. In fact, thirteen are from the Far East, four are from Europe (Western), two are from North America, and one is from the Middle East. Eighteen are from universities, and two are from research institutes. Total publications listed in the SCI for each of these twenty authors were scanned visually, and, on average, these authors were rarely listed as first authors. For example, in their 100 most recent papers, DAHN JR was listed as first author five times, and TARASCON JM was listed as first author six times.

4.1.1.2 Clustering Results

Appendix 1 presents three clustering approaches for generating groups of related technical areas. The base data is a square co-occurrence matrix of the highest frequency technical phrases. The two statistical clustering processes listed are used to cluster authors. A square co-occurrence matrix of the 253 most prolific authors (each matrix element represents the number of times each author pair is listed on the same paper) was generated, and clusters of related authors were then generated by the two statistical methods: factor matrix and multi-link aggregation.

4.1.1.2.1 Factor Matrix

The co-occurrence matrix was converted to a correlation matrix using the TechOasis software package. Then, the correlation matrix was converted to a factor matrix using the WINSTAT software package. Each matrix element contains the factor loading, a measure of the importance of each author to each factor. The number of factors was unconstrained, but the eigenvalues had a floor of unity. Practically, this means that each factor provides some additional useful information.

Appendix 2 contains the complete factor matrix for all 253 authors. Factor loading values above a threshold were shaded. Each column represents one factor, and the dark vertical bands in each column represent the essential contributors to each factor. The most cohesive factors start from the left column, and proceed to decrease in strength monotonically to the right. Thus, the strongest factor is the first, consisting of the authors ranging from Yokokawa to Yamaji. Further down the column, it is seen that Ishikawa is also related to the first factor, where Ishikawa is also in a separate factor with Morita and Matsuda. Since the factors are not fully orthogonal, and some authors can be connected strongly to more than one factor, all the authors that represent the core of a factor are not necessarily presented in contiguous form, as was the case with Ishikawa. Based on name ethnicity alone, intra-country clustering appears to be the dominant form of grouping. This will be examined later with country clustering.

4.1.1.2.2 Multi-Link Aggregation

To obtain a slightly different perspective on groupings, as well as to obtain more detail into sub-groupings within each factor, multi-link aggregation was performed. The co-occurrence matrix was normalized, and clusters of related authors were generated using the multi-link aggregation method of the WINSTAT software package described in Appendix 1.

Appendix 3 contains the full dendogram for the 253 most prolific authors. The ordinate is the ‘distance’, a measure of the coupling strength between authors, or groups of authors. Smaller ‘distance’ means stronger coupling. The abcissa is authors, and positioning of an author, or group of authors, along the axis also reflects the relationships between authors or between groups.

The authors from Factor 1 of the factor matrix are shown to constitute a cluster (close to the bottom of the dendogram). In this portrayal, however, the other factor with Ishikawa and Morita and Matsuda is shown as a cluster contiguous to the cluster identical to Factor 1. Additionally, the detailed structure within the Factor 1 cluster is evident from the dendogram. Yokohawa, Horita, and Sakai form a tightly knit unit; that unit in combination with Yamaji forms a moderately knit unit; and in combination with the moderately knit unit of Dokiya and Kawada, forms a less moderately knit. This less moderately knit unit is weakly coupled with the less moderately knit unit containing Ishikawa to form a larger cluster.

Table 2A is a co-occurrence matrix of the authors contained in Factor 1.

TABLE 2A - FACTOR 1 AUTHORS CO-OCCURRENCE MATRIX

# OF PAPERS	AUTHOR NAMES	YOKOKAWA	HORITA	SAKAI	DOKIYA	KAWADA	YAMAJI	ISHIKAWA
28	**YOKOKAWA**	28	25	25	16	14	15	5
25	**HORITA**	25	25	25	15	13	15	5
25	**SAKAI**	25	25	25	15	13	15	5
21	**DOKIYA**	16	15	15	21	13	5	3
18	**KAWADA**	14	13	13	13	18	4	1
15	**YAMAJI**	15	15	15	5	4	15	5
23	**ISHIKAWA**	5	5	5	3	1	5	23

The closeness of the sub-group members depicted schematically on the dendogram is confirmed by the actual numbers of papers. All of Horita'a and Sakai's papers are published with each other and with Yokohawa, and all three have very similar numbers of papers. Almost all of Dokiya's and Kawada's papers are published with each other and with the tightly knit first sub-group, but because their absolute numbers are somewhat lower than those of the first sub-group, they are placed in a separate sub-group, linked strongly to the first sub-group.

4.1.2 Journals Containing Most Electrochemical Power Papers

There were 587 different journals represented, with an average of 11.90 papers per journal. The journals containing the most power-related electrochemistry papers (see Table 3) had more than an order of magnitude more papers than the average.

TABLE 3 – JOURNALS CONTAINING MOST PAPERS

JOURNAL NAMES	# OF PAPERS
J. POWER SOURCES	1240
J. ELECTROCHEM. SOC.	771
SOLID STATE ION.	546
ELECTROCHIM. ACTA	403
J. ALLOY. COMPD.	290
DENKI KAGAKU	198
J. APPL. ELECTROCHEM.	167
J. ELECTROANAL. CHEM.	138
ELECTROCHEM. SOLID STATE LETT.	119
INT. J. HYDROG. ENERGY	112
RUSS. J. ELECTROCHEM.	100
ELECTROCHEMISTRY	86
J. MATER. CHEM.	81
J. SOLID STATE CHEM.	72
CHEM. MAT.	70
J. NEW MAT.ELECTROCHEM. SYST.	60
ELECTROCHEM. COMMUN.	56
SYNTH. MET.	55
BULL. ELECTROCHEM.	54
J. PHYS. CHEM. B	50

The majority of the journals are electrochemistry, with the remainder divided between chemistry and materials. There appear to be three primary groups at the top layer. The Journal of Power Sources, an international journal devoted to the science and technology of electrochemical energy systems, contains the most articles by far. This is not surprising, since its stated mission is fully aligned with the main objective of the present study. While many of its articles were retrieved by the query, essentially all of its articles are relevant to the topic of the present study.

The next group consists of the Journal of the Electrochemical Society (JES) and Solid State Ionics (SSI). The JES focuses on solid-state and electrochemical science and technology, while SSI is devoted to the physics, chemistry and materials science of diffusion, mass transport, and reactivity of solids. While these journals include aspects of electrochemistry/ electrochemical power sources in their charters, they include other aspects of chemistry (and physics) as well. The next five journals listed constitute the third group.

4.1.3 Institutions Producing Most Electrochemical Power Papers

A similar process was used to develop a frequency count of institutional address appearances. It should be noted that many different organizational components may be

included under the single organizational heading (e.g., Harvard Univ could include the Chemistry Department, Biology Department, Physics Department, etc.). Identifying the higher level institutions is instrumental for these DT studies. Once they have been identified through bibliometric analysis, subsequent measures may be taken (if desired) to identify particular departments within an institution.

TABLE 4 – PROLIFIC INSTITUTIONS

INSTITUTION NAMES	COUNTRY	# OF PAPERS
CHINESE ACAD SCI	CHINA	118
KYOTO UNIV	JAPAN	108
CNRS	FRANCE	104
KOREA ADV INST SCI & TECHNOL	KOREA	90
RUSSIAN ACAD SCI	RUSSIA	89
ZHEJIANG UNIV	CHINA	85
ARGONNE NATL LAB	USA	79
UNIV CALIF BERKELEY	USA	78
TOHOKU UNIV	JAPAN	73
MIT	USA	66
CNR	ITALY	63
CENT ELECTROCHEM RES INST	INDIA	60
SEOUL NATL UNIV	KOREA	60
TOKYO INST TECHNOL	JAPAN	55
CSIC	SPAIN	55
KFA JULICH GMBH	GERMANY	54
UNIV S CAROLINA	USA	54
OSAKA NATL RES INST	JAPAN	52
UNIV TOKYO	JAPAN	51
DELFT UNIV TECHNOL	NETHERLANDS	51

Of the twenty most prolific institutions, ten are from Asia, five are from Western Europe, four from the USA, and one from Eastern Europe. Twelve are universities, and the remaining institutions are research institutes.

4.1.4 Countries Producing Most Electrochemical Power Papers

There are 78 different countries listed in the results. The country bibliometric results are summarized tabularly in Table 5 and graphically in Figure 1. The dominance of a handful of countries is clearly evident.

TABLE 5 – PROLIFIC COUNTRIES

COUNTRY NAMES	# OF PAPERS
JAPAN	1552
USA	1318
FRANCE	558
PEOPLES R CHINA	499
SOUTH KOREA	380
GERMANY	341
CANADA	318
ENGLAND	285
ITALY	250
INDIA	249
RUSSIA	206
SPAIN	151
SWEDEN	126
AUSTRALIA	121
SWITZERLAND	113
NETHERLANDS	97
TAIWAN	90
BRAZIL	83
ISRAEL	78
POLAND	73

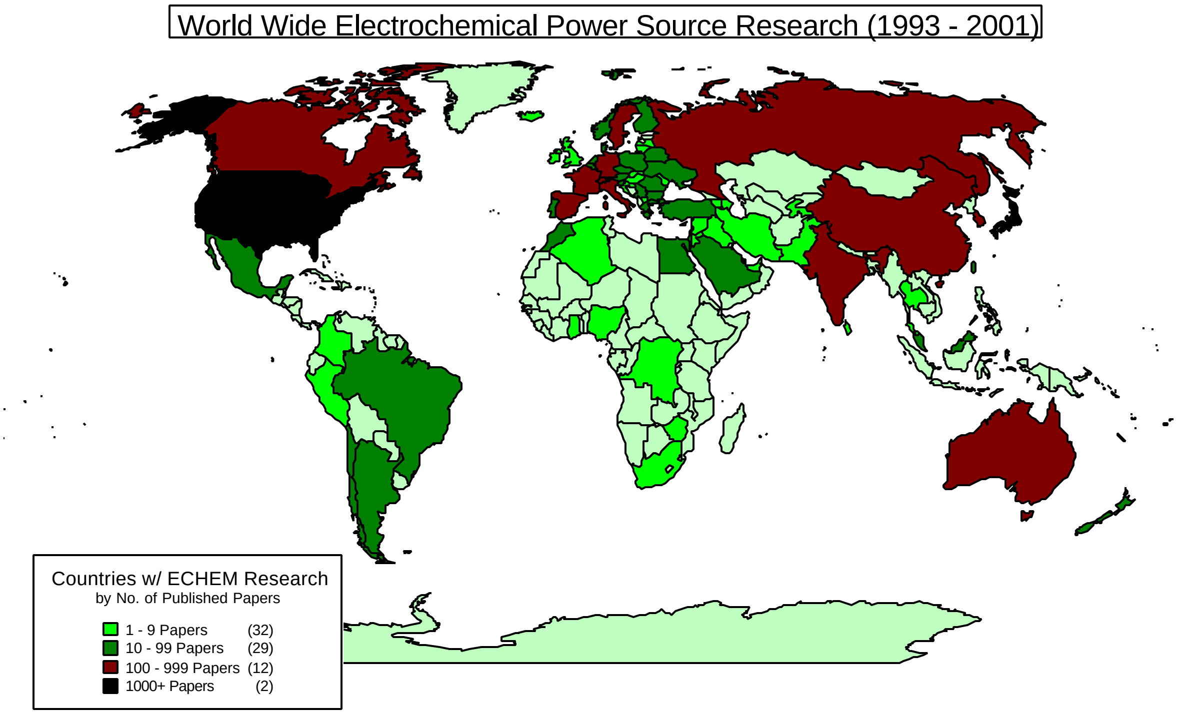


FIGURE 1 – COUNTRY BIBLIOMETRIC RESULTS

There appear to be three dominant groups in the twenty most prolific countries. The US and Japan constitute the most dominant group, and were the only two countries to have published more than 1000 papers on power-related electrochemistry during the past 8 years. France and China constitute the next group, but had less papers combined than either member of the first group. The next seven countries constitute the third group.

Interestingly, unlike all previous DT studies, the United States (US) was not the most prolific country. Japan had more published papers (nearly 18% more) than the US. Overall, Eastern Asia (Japan, China, South Korea, Taiwan), Northern North America (US, Canada), and Western Europe (France, Germany, UK) accounted for most of the electrochemistry research activity. Other world regions have also made contributions (Figure 2).

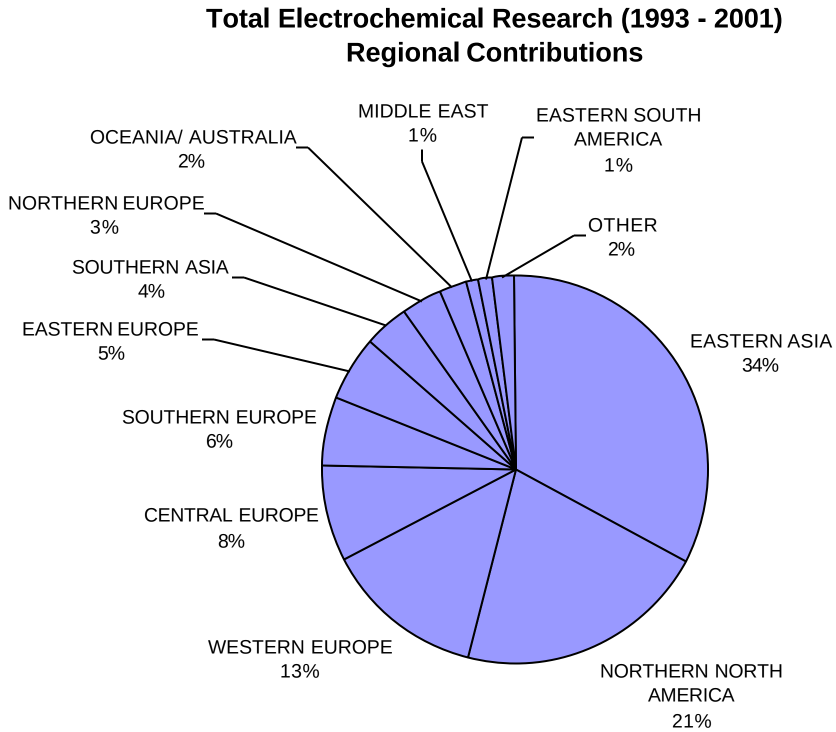


FIGURE 2 – REGIONAL CONTRIBUTIONS

Appendix 4 contains a co-occurrence matrix of the top 15 countries. In terms of absolute numbers of co-authored papers, the USA major partners are Japan, France, Italy, Canada, and South Korea. Overall, countries in similar geographical regions tend to co-publish substantially, the US being a moderate exception.

4.2 Citation Statistics on Authors, Papers, and Journals

The second group of metrics presented is counts of citations to papers published by different entities. While citations are ordinarily used as impact or quality metrics (14), much caution needs to be exercised in their frequency count interpretation, since there are numerous reasons why authors cite or do not cite particular papers (15, 16).

The citations in all the retrieved SCI papers were aggregated, the authors, specific papers, years, journals, and countries cited most frequently were identified, and were presented in order of decreasing frequency. A small percentage of any of these categories received large numbers of citations. From the citation year results, the most recent papers tended to be the most highly cited. This reflected rapidly evolving fields of research.

4.2.1 Most Cited Authors

The most highly cited authors are listed in Table 6.

TABLE 6 – MOST CITED AUTHORS
(cited by other papers in this database only)

AUTHOR NAMES	INSTITUTIONS	COUNTRIES	TIMES CITED
OHZUKU, T	OSAKA CITY UNIV	JAPAN	1066
THACKERAY, MM	ARGONNE NAT'L LAB	USA	845
AURBACH, D	BAR ILAN UNIV	ISRAEL	808
TARASCON, JM	UNIV PICARDIE	FRANCE	755
DAHN, JR	DALHOUSIE UNIV	CANADA	698
WATANABE, M	UNIV YAMANASHI	JAPAN	601
ABRAHAM, KM	COVALENT ASSOCIATES	USA	461
GUMMOW, RJ	CSIR	SOUTH AFRICA	455
DELMAS, C	CNRS	FRANCE	429
SAKAI, T	OSAKA NAT'L RES INST	JAPAN	412
PISTOIA, G	CNR	ITALY	391
MINH, NQ	ALLIED SIGNAL AERO	USA	381
GOODENOUGH, JB	UNIV TEXAS	USA	379
ISHIHARA, T	OITA UNIV	JAPAN	370
STEELE, BCH	UNIV LONDON IMPERIAL	ENGLAND	351
REIMERS, JN	MOLI ENERGY	CANADA	345
PELED, E	TEL AVIV UNIV	ISRAEL	335
GUYOMARD, D	UNIV NANTES	FRANCE	332
MIZUSAKI, J	TOHOKU UNIV	JAPAN	324
APPLEBY, AJ	TEXAS A&M	USA	300

Of the twenty most cited authors, five are from Japan, five from the USA, five from Europe (Western), two from Canada, two from Israel, and one from Africa. This is a far different distribution from the most prolific authors, where thirteen were from the Far East. There are a number of potential reasons for this difference, including difference in quality and late entry into the research discipline. In another three or four years, when the papers from present-day authors have accumulated sufficient citations, firmer conclusions about quality can be drawn.

The lists of twenty most prolific authors and twenty most highly cited authors only had four names in common (AURBACH, TARASCON, DAHN, WATANABE). This phenomenon of minimal intersection has been observed in all other text mining studies performed by the first author.

Thirteen of the authors' institutions are universities, four are government-sponsored research laboratories, and three are private companies. The appearance of the companies on this list is another differentiator from the list of most prolific authors.

The citation data for authors and journals represents citations generated only by the specific records extracted from the SCI database for this study. It does not represent all the citations received by the references in those records; these references in the database records could have been cited additionally by papers in other technical disciplines.

4.2.2 Most Cited Papers

The most highly cited papers are listed in Table 7.

TABLE 7 – MOST CITED PAPERS
(total citations listed in SCI)

AUTHOR NAME	YEAR	JOURNAL	VOLUME	SCI CITES
TARASCON JM	1991	J ELECTROCHEM SOC	V138	272
(LIMN204 SPINEL PHASE AS SECONDARY LITHIUM CELL CATHODE)				
MINH NQ	1993	J AM CERAM SOC	V76	476
(CERAMIC FUEL CELLS - REVIEW)				
OHZUKU T	1993	J ELECTROCHEM SOC	V140	217
(SYNTHESIS OF LINIO2 FOR SECONDARY LITHIUM CELL)				
GUMMOW RJ	1994	SOLID STATE IONICS	V69	281
(IMPROVED RECHARGEABLE CAPACITY OF LIMN2O4 CATHODES)				
OHZUKU T	1990	J ELECTROCHEM SOC	V137	314
(ELECTROCHEMISTRY OF MNO2 IN LITHIUM CELLS)				
MIZUSHIMA K	1980	MATER RES BULL	V15	392
(LIXCOO2 FOR HIGH-ENERGY DENSITY BATTERY CATHODES)				
GUYOMARD D	1992	J ELECTROCHEM SOC	V139	300
(LI METAL-FREE RECHARGEABLE LIMN2O4/ CARBON CELLS)				
THACKERAY MM	1983	MATER RES BULL	V18	358
(LITHIUM INSERTION INTO MANGANESE SPINELS)				
TARASCON JM	1994	J ELECTROCHEM SOC	V141	247
(LITHIUM INSERTION INTO THE SPINEL LIMN2O4)				
FONG R	1990	J ELECTROCHEM SOC	V137	334
(LITHIUM INTERCALATION INTO CARBON USING NON-AQUEOUS CELLS)				
REIMERS JN	1992	J ELECTROCHEM SOC	V139	227
(LITHIUM INTERCALATION IN LIXCOO2)				
COURTNEY IA	1997	J ELECTROCHEM SOC	V144	147
(LITHIUM REACTION WITH TIN OXIDE COMPOSITES IN LITHIUM ION CELL)				
SATO K	1994	SCIENCE	V254	221
(LITHIUM STORAGE IN DISORDERED CARBONS)				
THACKERAY MM	1992	J ELECTROCHEM SOC	V139	202
(SPINEL ELECTRODES FROM LIMNO SYSTEM FOR SECONDARY BATTERIES)				
THACKERAY MM	1984	MATER RES BULL	V19	235
(ELECTROCHEMICAL EXTRACTION OF LITHIUM FROM LIMN2O4)				

ISHIHARA T	1994 J AMER CHEM SOC	V116	201
(DOPED LAGO3 OEROVSKITE OXIDE IONIC CONDUCTOR)			
SHANNON RD	1976 ACTA CRYSTALLOGR A	V32	10254
(IONIC-RADII AND INTERATOMIC DISTANCES IN HALIDES AND CHALCOGENIDES)			
WILLEMS JJG	1984 PHILLIPS J RESEARCH	V39	285
(METAL HYDRIDE ELECTRODES FOR RECHARGEABLE BATTERY)			
ABRAHAM KM	1990 J ELECTROCHEM SOC	V137	202
(LI+-CONDUCTIVE SOLID POLYMER ELECTROLYTES WITH LIQ-LIKE CONDUCT)			
OHZUKU T	1993 ELECTROCHIMICA ACTA	V38	139
(LI-N-CO OXIDES FOR SECONDARY LITHIUM CELLS			

The theme of each paper is shown in italics on the line after the paper listing. The order of paper listings is by number of citations by other papers in the extracted database analyzed. The total number of citations from the SCI paper listing, a more accurate measure of total impact, is shown in the last column on the right.

The Journal of the Electrochemical Society contains the most papers by far, twelve out of the twenty listed. Most of the journals are fundamental science journals, and most of the topics have a fundamental science theme. Most of the papers are from the 1990s, with four being from the 1980s, and one extremely highly cited paper being from 1976. This reflects a dynamic research field, with seminal works being performed in the recent past.

Sixteen of the papers address issues related to lithium secondary batteries, with the dominant issue theme being lithium insertion/ intercalation to avoid free-metal formation. Two of the papers address issues related to ceramic fuel cells, with the dominant issue theme being solid oxides for high ionic conductivity. One paper addresses issues related to nickel metal hydride rechargeable batteries.

Thus, the major intellectual emphasis of cutting edge electrochemical power sources research, as evidenced by the most cited papers, is well aligned with the intellectual heritage and performance emphasis, as will be evidenced by the clustering approaches.

4.2.3. Most Cited Journals

TABLE 8 – MOST CITED JOURNALS
(cited by other papers in this database only)

JOURNAL NAMES	TIMES CITED
J ELECTROCHEM SOC	22363
SOLID STATE IONICS	9782
J POWER SOURCES	8265
ELECTROCHIM ACTA	5994
J ELECTROANAL CHEM	4607
J SOLID STATE CHEM	2364
J ALLOY COMPD	2269
J APPL ELECTROCHEM	2008
MATER RES BULL	1811
PHYS REV B	1672
J AM CHEM SOC	1491
J PHYS CHEM-US	1470
J AM CERAM SOC	1417
J LESS-COMMON MET	1399
DENKI KAGAKU	1157
SYNTHETIC MET	1041
CHEM MATER	969
ELECTROCHEMICAL SOC	851
SCIENCE	841

The Journal of the Electrochemical Society received as many citations as the next three journals combined. Most of the highly cited journals are electrochemistry, some are materials, some chemistry, with one physics journal represented. Based on all the citation results, there is little evidence that disciplines outside the tightly knit electrochemistry-materials groups relevant to the specific applications are being accessed.

The authors end this bibliometrics section by recommending that the reader interested in researching the topical field of interest would be well-advised to, first, obtain the highly-cited papers listed and, second, peruse those sources that are highly cited and/or contain large numbers of recently published papers.

4.3 Database Tomography Results

There are two major analytic methods used in this section to generate taxonomies of the SCI databases: non-statistical clustering, based on phrase frequency analysis, and statistical clustering, based on phrase proximity analysis. They are described in Appendix 1.

4.3.1. Taxonomies

4.3.1.1 Non-Statistical Clustering

4.3.1.1.1. Keyword Taxonomy

All the Keywords from the extracted SCI records, and their associated frequencies of occurrence, were tabulated, and then grouped into categories by visual inspection. The phrases were of two types: system-related and tech base-related. While the system sub-categories were relatively independent, there was substantial overlap between some of the tech base categories. In particular, the generic system components (electrolytes, electrodes) had overlap with the generic tech base components. The detailed taxonomy results are presented in Appendix 5. Within each sub-category, thrust areas are presented in approximate frequency appearance order. These results are summarized now.

The main system categories were Sources-Converters, Storage, and Other Applications. Sources-Converters included Fuel Cells, Storage included Batteries and Capacitors (to a much smaller extent), and Other Applications covered many uses of Electrochemical Cells.

The main tech-base categories were Electrolytes, Electrodes, Materials, Processes, Properties, Phenomena, Parameters, Experiments, Theory, Macrostructure, Microstructure, and Region.

Fuel cell concepts emphasized were solid oxide, molten carbonate, and polymer electrolyte. While most efforts addressed pure hydrogen input, or reformed hydrocarbon input to supply hydrogen (e.g., the CO-tolerant internal reforming molten carbonate fuel cell), some fuel cell research efforts focused on direct hydrocarbon conversion. These concepts included e.g. solid metal electrolyte and polymer electrolyte fuel cells.

Batteries focused strongly on lithium-ion rechargeable, with substantial effort on nickel-metal hydride and lead-acid, and somewhat less emphasis on rechargeable air-electrode batteries. The lithium-ion battery emphasis was two-fold: increasing electrode intercalation to minimize lithium free-metal formation with non-aqueous electrolytes, and modifying the electrolyte (and electrodes) to increase conductivity and allow higher specific energies. Research emphasis to improve specific energy appeared to be on solid-polymer electrolytes, backed by highly conducting current collectors. The nickel-metal hydride battery is a potential replacement for the balanced performance (specific energy, specific power, cycle-life, reliability) but environmentally contaminating Ni-Cd battery. Ni-M-H emphasized developing metal hydrides for negative electrodes that can be decomposed and reformed reversibly, and have high hydrogen retention to reduce battery self-discharge rate. The lead-acid battery emphasis appears to be materials, packaging, and fabrication. Much of the associated modeling and simulation

reflected in the Keywords was related to lead-acid battery system optimization, focusing on maximizing energy at optimum power density, minimizing internal resistance, retaining maximum charge, and maximizing mechanical strength and cycle life. Environmental impact of lead-acid batteries, mainly disposal and material recyclability, was a research area as well. The rechargeable air-electrode batteries, such as zinc-air, focused on methods to increase charging rates and reduce electrolyte carbonation from air-based carbon compounds.

There was very little effort reflected in capacitors. Relatively high specific power/ specific energy super-capacitors appear to be a key research area, with potential for filling the gap between the high power density low energy density conventional capacitors and the high energy density low power batteries or very high energy density low power density fuel cells. Research focus appears aimed at electrode geometries (larger surface areas) and materials that will allow higher capacitance, and electrolyte materials (and geometries) that will allow the relatively low voltage ceilings (due to potential breakdown across the small separator gap) to be raised.

4.3.1.1.2 Abstract Taxonomy

A taxonomy of electrochemical energy-related technologies was developed through visual inspection of the Abstract phrase frequencies. The developed taxonomy was subsequently used to approximate global levels of emphasis (GLE). This type of analysis would help identify adequately and inadequately supported system and subsystem tech base areas. It could also differentiate the developed and developing technology components of a particular system.

In this section, a three level taxonomy was required to provide sufficient detail on the various electrochemical energy-related technologies. The first two levels of the taxonomy were developed using a phrase frequency-only analysis. Phrases generated with the phrase frequency analysis could be classified into two generic types of categories: system specific (e.g., SOLID OXIDE FUEL CELL, LITHIUM ION BATTERIES, STEAM REFORMING, ELECTRIC DOUBLE-LAYER CAPACITORS) and generic (ELECTROLYTES, CAPACITY, ELECTRODES, DISCHARGE, CATHODES, ANODES). Since one feature of the manually generated taxonomy was allocation of Abstract phrases and associated frequencies to specific categories in order to estimate GLE of specific systems, a method was required to relate the generic phrases to their associated specific systems (e.g., what fraction of the ELECTRODES frequencies should be allocated to the BATTERIES or FUEL CELLS categories?).

The method selected was to perform a proximity analysis using the second level taxonomy categories as themes. The second level of the taxonomy consisted of high technical content system-specific phrases from the phrase frequency analysis data. Phrases in close physical (and thematic) proximity to the system-specific phrases were generated, and the more generic tech base phrases were assigned to the related system-specific categories weighted by their occurrence frequencies.

The results of the taxonomy development and GLE analyses are presented in Table A6-1 of Appendix 6. The sum of the phrase frequencies for each category, used as the proxy metric for the GLE, is given in parentheses, immediately following the taxonomy group heading.

4.3.1.1.2.1 Taxonomy Level 1

The highest taxonomy level consisted of two categories: Electrochemical Converters (17,227) that were comprised of fuel cell technologies, and Electrochemical Source and Storage Devices (24,804) consisted of battery and electrochemical capacitor technologies.

An evaluation of the current status of these technologies showed that electrochemical source/storage devices provided power to a wider variety of applications (especially for small / portable electronic devices) compared to fuel cells. As the numbers of applications requiring portable power continue to increase rapidly, near term solutions are required, and batteries currently seem to offer the most feasible solution. Hence, the higher level of emphasis in that category.

The literature indicated that Fuel Cell research was substantial. The potential of fuel cell technology is promising, but many technical and economic issues such as the need for expensive catalysts (in low temperature fuel cells), the large size / weight / miniaturization, and the corrosion / breakdown of components (for high temperature Fuel Cells) would need to be resolved for it to become more competitive. Another issue is the lack of infrastructure for some of the fuels (methanol, ethanol, and hydrogen) required to supply these Fuel Cells.

Supercapacitors remain ideal for delivering high power over a short duration but are most useful when combined with other electric power sources (i.e. to provide extra power boost in an electric application). Technical advances were still required before higher energy density systems could be mass-produced.

4.3.1.1.3 Taxonomy Level 2

4.3.1.1.3.1 Fuel Cells

Fuel Cell research addressed the following tech base areas: system components and component configurations (4,038); properties and characteristics (2,683); sources / fuels (2,385); materials (2,358); conversion processes (1,363); conversion byproducts (1,011); operating conditions (885); and potential applications (267).

In general, Fuel Cell research was aimed at improving the performance of Fuel Cells (for a wider variety of applications) while lowering the cost (manufacturing, operation, and maintenance) of the systems. Advances in the individual Fuel Cell components could help achieve many of those objectives. For example:

- Use of porous electrodes to increase electrode surface area, and subsequently, improve electrode current density and overall fuel cell performance;
- Use of polymer electrolytes to reduce costly corrosion and electrolyte loss problems, and to improve hydrogen ion conductivity (which improves fuel cell performance);
- Development of electrocatalysts to improve the rate of electrode reactions / exchange current density.

The importance of advances in component technologies was reflected by the high GLE for that category.

The Fuel Cell properties / characteristics (such as ionic conductivity, current density, corrosion rate), sources / fuels (such as hydrogen, methane, natural gas, hydrocarbons, methanol, ethanol), and materials (such as nafion, zironia, platinum, ceramics) seemed to have equal amounts of research according to the GLE. All three of these categories provided critical supporting / enabling technologies for improving Fuel Cells from a performance or cost perspective.

Other important Fuel Cell categories were conversion processes, byproducts, and operating conditions. Lastly, there was little emphasis on applications. Current Fuel Cell technologies did not address a high variety of applications (based on range of required energy or rate of power output)..

Most of the recent research on Fuel Cell technologies focused on solid oxide and molten carbonate Fuel Cells, due to their high operating temperatures and consequently higher efficiencies and relaxed reforming requirements . Polymer electrolyte / proton exchange membrane (PEM) Fuel Cells were also being pursued, but at a lower level. The Fuel Cells' main technical areas of interest were the same as described in the Keyword taxonomy. For solid oxide Fuel Cells (SOFC) electrolyte doping, synthesis and characterization of various anode / cathode materials, and the development of low temperature SOFC seemed to be the focal points of research. Molten carbonate fuel cell (MCFC) research was focused on the development of novel cathode materials (often a nickel based alloy for corrosion resistance between the cathode and current collector plate), and the selection of anode catalyst for the reforming reaction. Polymer electrolyte and Proton Exchange Membrane (PEM) Fuel Cell research was developing new techniques for preparing the catalyst layer in the electrolyte, developing new fabrication methods for composite membranes, and was assessing the physical and morphological characteristics and electrochemical behavior of various catalysts (i.e. PtRu/C).

4.3.1.1.3.2 Batteries

Batteries research addressed the following tech base areas: materials (7,850), properties and characteristics (4,643), component technologies (4,531), processes and phenomena (2,658), types (2,195), and applications (1,121)

Materials seemed to be an integral part of advancing battery technologies. The enhanced properties and characteristics of novel materials, and their application in the electrodes, electrolytes, or at the interface were projected to result in lighter, higher energy capacity, less expensive batteries. For example, research was being conducted on:

- porous metals / high surface area materials for NiCd / NiMH electrode;
- hydrogen absorbing metals for NiMH anodes;
- alloy powders for higher energy capacity and improved cycle life;
- carbon coated silicon for anodes to improved cycleability;
- analyses of reactions at electrode-electrolyte (non-aqueous) interface;
- polymer electrolytes;
- carbon anodes / composite anodes;
- microwave synthesis of electrode materials;

Materials were expected to improve properties and characteristics of batteries and battery components (electrodes and electrolytes) with similar impacts. This was reflected by the equal amounts of emphasis in each category, according to the GLE numbers.

Little was discussed in the literature about the basic theory behind the electrochemical processes on which batteries are based, other than material synthesis and characterization. There were some efforts in the areas of modeling and simulation, mainly focused on lead-acid batteries. There was a modest amount of the literature that focused on applications of batteries. The reputation, variety, and current availability of batteries has made them the top option for many applications requiring portable electric power.

Technical area research for battery technologies had a strong focus on secondary lithium batteries (Lithium Ion, Lithium Polymer, etc.). Research included the investigation of carbon-carbon (C-C) composite as an anode material for lithium ion batteries. Lithium-ion cells made with the C-C composite anode showed many advantages, such as excellent performance and enhanced safety. Other research focused on improving the charge-discharge characteristics of polyaniline films used as cathodes in lithium batteries and improving the ionic conductivity of crystalline polymer electrolytes. Lead-acid batteries also continued to be a major thrust area. Research there focused on battery additives and their influence on the separator behavior. There was research on the use of gas-recombining noble metal catalysts in valve regulated lead acid batteries. Also, researchers were developing high rate discharge (HRD) lead-acid batteries, geared for automobile applications. Nickel Metal Hydride (NiMH) batteries constituted another major thrust area. Research was conducted to develop new / improved methods for synthesizing and characterizing materials. For example, surface modification of hydrogen storage alloy electrodes by the hot-charging treatment was being investigated. New hydrogen storage materials were being developed. Gelatin-pretreated graphite anodes were being tested to see if they could improve the irreversible loss of charge, reversible capacity and efficiency of NiMH batteries.

4.3.1.1.3.3 Electrochemical Capacitors

Electrochemical capacitors research addressed the following tech base areas: Properties and characteristics (604); Component technologies (568); materials (435); and types (199). The major concerns were associated with improving the energy density, and power density, hence the higher GLE for that category. Potential solutions for achieving this included improving components (dielectric / gel electrolyte, solid electrolyte; polarizable electrodes/ composite electrodes) and improvements based on component materials (glassy carbon, carbon fibers, aerogels, thin films).

A focal point of electrochemical capacitor research was the chemical synthesis and characterization of various materials (i.e. MoxSy (CO)(n) and Mo-x(CO)(n)) and their possible application as catalysts for oxygen reduction reaction. Other electrochemical capacitor thrust areas included: the use of industrial carbon blacks (CBs) materials in electrochemical supercapacitors because of their high specific areas; the development of thin film supercapacitors using a sputtered RuO2 electrode; the development of carbon nanotubes/RuO2 electrodes for electrochemical capacitors; the synthesis, structure-property characterization, and performance of carbon aerogels; and the fabrication and application of Cu-carbon composite (prepared from sawdust) to electrochemical capacitor electrodes.

The absence of any categories/ sub-categories in this taxonomy should not be interpreted that S&T efforts are not being pursued in those areas. The correct interpretation is that within the frequency threshold constraints of the electrochemical database, mid-high frequency phrases related to these categories do not appear.

4.3.1.2. Statistical Clustering

Two statistically-based clustering methods were used to develop taxonomies, factor matrix clustering and multi-link clustering. Both offer different perspectives on taxonomy category structure from the non-statistical manual clustering approach described above. None of the three approaches are inherently superior.

Appendix 1 describes the statistically-based clustering methodologies in more detail. Appendix 1A overviews the generic statistically-based clustering approach, Appendix 1B describes the factor matrix clustering, and Appendix 1C describes the multi-link clustering.

4.3.1.2.1. Factor Matrix Clustering

A correlation matrix of the 218 highest frequency high technical content phrases was generated, and a factor analysis was performed using the WINSTAT statistical package. The eigenvalue floor was set equal to unity, and a factor matrix consisting of 20 factors resulted. A description of these factors, and their aggregation into a taxonomy, follows.

The capitalized phrases in parentheses represent typical high factor loading phrases for that factor. The complete factor matrix is presented in Appendix 7.

Factor 1 (CR, MN, FE, NI, TI, V, ZR, CU, MG, MO, HYDRIDE, DISCHARGE CAPACITY, ELECTROCHEMICAL PROPERTIES) – materials used in electrochemical systems to improve discharge capacity, especially for cathodes; some emphasis on hydrogen storage alloys for metal hydride batteries.

Factor 2 (ELECTRICAL CONDUCTIVITY, IONIC CONDUCTIVITY, SR, OXYGEN PARTIAL PRESSURE, THERMAL EXPANSION, SOLID OXIDE FUEL CELL) – focuses on rare earth oxides for higher electrical and ionic conductivity and lower thermal expansion electrodes in solid oxide fuel cells.

Factor 3 (METHANOL OXIDATION, PLATINUM, RUTHENIUM, CATALYSTS) – focuses on the use of platinum/ ruthenium catalysts for enhancing methanol oxidation at electrodes of direct methanol oxidation fuel cells.

Factor 4 (LA, ND, PR, YSZ, SOLID OXIDE FUEL CELLS, THERMAL EXPANSION) – focuses on rare earth oxide composite electrodes, especially with YSZ added to either/ both electrolytes and electrodes, to increase conductivity and reduce thermal expansion in solid oxide fuel cells.

Factor 5 (LITHIUM, INTERCALATION, ELECTROCHEMICAL INTERCALATION, CHEMICAL DIFFUSION COEFFICIENT, GRAPHITE, SECONDARY LITHIUM BATTERIES) – focuses on intercalation of lithium into graphite-based electrodes, as evidenced by chemical diffusion coefficient and other metrics, for lithium ion secondary batteries.

Factor 6 (ANODE, CATHODE, CURRENT DENSITY, MAXIMUM POWER DENSITY, FUEL CELL, SOFC) – focuses on fuel cell performance characteristics (especially SOFC), such as current density and maximum power density, as functions of different electrode materials.

Factor 7 (PROPYLENE CARBONATE, ETHYLENE CARBONATE, ELECTROLYTES, LI, IONIC CONDUCTIVITY, GRAPHITE) – focuses on interactions between lithium-based electrodes and non-aqueous carbonate-based electrolytes for lithium-ion secondary batteries.

Factor 8 (NA, K, LI, ELECTROCHEMICAL INTERCALATION, S) – focuses on electrochemical intercalation of sodium family-based compounds, especially lithium, especially in graphite, especially for re-chargeable lithium-ion batteries, and the role of sulfur compound additives in enhancing the intercalation-deintercalation process.

Factor 9 (CATALYSTS, OXYGEN REDUCTION, CARBON BLACK, PLATINUM, NITROGEN, METHANOL) – focuses on platinum-based catalysts on carbon-based electrodes for enhancing oxygen reduction in fuel cells.

Factor 10 (HYDROGEN, METHANE, FUEL CELL, STORAGE, ETHANOL, METHANOL, CATALYST, MOLTEN CARBONATE FUEL CELLS) – focuses on conversion of methane (or methanol/ ethanol) to hydrogen for use in fuel cells, especially molten carbonate fuel cells, and addresses storage of hydrogen in solid material in parallel.

Factor 11 (THIN FILMS, CYCLIC VOLTAMMETRY, CHRONOAMPEROMETRY, IMPEDANCE SPECTROSCOPY, LITHIUM IONS) – focuses on electrochemical characterization of thin film electrodes, including intercalation behaviour, mainly for lithium secondary batteries.

Factor 12 (SPINEL STRUCTURE, CAPACITY FADING, SOL-GEL METHOD, CATHODE MATERIAL, CYCLABILITY, RECHARGEABLE LITHIUM BATTERIES) – focuses on use of sol-gel method to fabricate spinel lithium compounds, and subsequent capacity fading and cyclability as a function of sol-gel materials and environmental parameters, for eventual use in lithium secondary batteries.

Factor 13 (NICKEL, IRON, COBALT, TITANIUM, COPPER, MAGNESIUM, MANGANESE, ALUMINUM, REVERSIBILITY, ELECTROCHEMICAL BEHAVIOUR, SPECIFIC CAPACITY, BATTERIES) – materials for electrochemical systems, especially for electrodes, especially for batteries.

Factor 14 (X-RAY DIFFRACTION, TRANSMISSION ELECTRON MICROSCOPY, X-RAY PHOTOELECTRON SPECTROSCOPY, ELECTROCHEMICAL MEASUREMENTS, THERMOGRAVIMETRIC ANALYSIS, REVERSIBLE CAPACITY, FILMS, CRYSTAL STRUCTURE) – experimental measurements of structures and properties and reactive phenomena to characterize electrochemical systems.

Factor 15 (DISCHARGE, CHARGE, LEAD-ACID BATTERIES, ELECTROCHEMICAL REACTIONS, POWER, SIMULATIONS, MATHEMATICAL MODEL, CORROSION) – focuses on lead-acid battery phenomena, such as charging and discharging, electrochemical reactions and subsequent corrosion, including modeling and simulation of battery processes.

Factor 16 (PROPYLENE CARBONATE, ETHELYNE CARBONATE, LITHIUM SALTS, POLYMER ELECTROLYTE, IONIC CONDUCTIVITY) – focuses on increasing electrolyte electrical and ionic conductivities by varying composition and concentrations of polymer, lithium salt, and plasticizer in polymer electrolytes, emphasizing propylene carbonate and ethylene carbonate plasticizers, for eventual use in secondary lithium-ion batteries.

Factor 17 (ELECTRODE SURFACE, HYDROGEN DIFFUSION, EXCHANGE CURRENT DENSITY, CYCLIC VOLTAMMETRY, ELECTRODE PERFORMANCE, ELECTROCHEMICAL IMPEDANCE SPECTROSCOPY, CATHODIC POLARIZATION, DISCHARGE CAPACITY) – focuses on measurement of hydrogen diffusion and charge

transfer (exchange current density) across electrodes, mainly for their effect on discharge capacity, mainly for metal hydrides as a precursor to NI-M-H batteries.

Factor 18 (NEGATIVE ELECTRODE, POSITIVE ELECTRODE, LITHIUM METAL, MANGANESE, COBALT, RECHARGEABLE LITHIUM BATTERIES, SPECIFIC CAPACITY, LEAD-ACID BATTERY) – focuses on use of layered lithium oxides, especially cobalt and manganese, as cathode materials that can reversibly intercalate lithium ions for lithium-ion rechargeable batteries.

Factor 19 (ENERGY DENSITY, CAPACITOR, POWER DENSITY, SPECIFIC ENERGY, CAPACITANCE, BATTERY, ELECTRIC VEHICLES, CYCLE LIFE) – primary focus on advanced capacitors with reasonable energy density for high power density high discharge long cycle life applications, with secondary focus on hybrid battery/ capacitor systems for high energy density high power density applications.

Factor 20 (REVERSIBLE CAPACITY, GRAPHITE, CARBON, TIN, ANODE MATERIAL, LITHIUM, LITHIUM-ION BATTERIES, FIRST DISCHARGE) – focuses on use of graphite-tin composites as anode materials for high reversible capacity lithium-ion batteries.

Thus, the 20 factors can be viewed as thrust areas constituting the lowest level taxonomy. Each factor contains one or more of the following elements: 1) System-specific (e.g., lithium-ion battery); 2) System-generic (e.g., re-chargeable battery); 3) Class (e.g., battery); 4) Phenomenon (e.g., electrolyte conductivity). Thus, there are myriad ways to combine the factors, depending on which dominant characteristics are chosen. In practice, the aggregation methodology will depend on the application for the taxonomy. For example, if the taxonomy is used to identify participants for a comprehensive workshop on electrochemical power sources, categorizing by phenomena would identify e.g., intercalation experts, while categorizing by system would identify e.g. lithium-ion battery experts. Selection of aggregation attributes would depend on the workshop objectives. Conversely, assume the taxonomy is used by a program manager to estimate global levels of effort in specific technologies, in order to identify technology areas of adequacy and deficiency. Then, categorizing by phenomena would identify e.g., intercalation deficiencies, whereas categorizing by system would identify e.g., lithium-ion battery deficiencies.

The factors above were aggregated by generic system at the highest level, then by specific system (when identifiable) at the next level, or by phenomenon when specific system is not identifiable, then the specific systems were de-aggregated to phenomena at the following level. The following hierarchical level taxonomy resulted (numbers in parenthesis are factor numbers from above).

The highest level taxonomy (level 1) consisted of Fuel Cells (2, 3, 4, 6, 9, 10), Batteries (5, 7, 8, 11, 12, 15, 16, 17, 18, 20), and Capacitors (19). The next highest level taxonomy (level 2) is:

Fuel Cells
*SOFC (2, 4, 6) – improve electric conductivity and thin film properties, and reduce thermal expansion, to increase current density and maximum power density.
*DMFC (3, 9, 10) – catalysts and hydrogen storage alloys to improve hydrocarbon oxidation and hydrogen storage.
*Materials (1, 13) – materials for electrochemical systems, including fuel cells.
*Diagnostics (14) – diagnostic techniques used to characterize phenomena and properties of electrochemical systems, including fuel cells.

Batteries
*Lithium-ion (5, 7, 8, 11, 12, 16, 18, 20) – rechargeable.
-Intercalation (5, 8, 11, 18, 20)
-Thin films (11)
-Conductivity (7, 16)
-Reversibility (12, 18, 20)
-Fabrication (12)
*Lead-Acid (15) – modeling. simulation, and performance characteristics measurement.
*Ni-M-H (1, 17) – hydrogen storage alloys structure and reactions.
*Materials (1, 13) – materials for electrochemical systems, including batteries.
*Diagnostics (14) – diagnostic techniques used to characterize phenomena and properties of electrochemical systems, including batteries.

Capacitors
*Thin Films (19) – capacitance and energy density of thin-film capacitors.

4.3.1.2.2. Multi-Link Clustering

A symmetrical co-occurrence matrix of the 217 highest frequency high technical content phrases was generated. The matrix elements were normalized using the Equivalence Index (Eij=Cij^2/Ci*Cj, where Ci is the total occurrence frequency of the ith phrase, and Cj is the total occurrence frequency of the jth phrase, for the matrix element ij), and a multi-link clustering analysis was performed using the WINSTAT statistical package. The Average Linkage method was used. Three types of raw data output were generated by each clustering run: a dendogram, a table, and a taxonomy. These three types of data output are described in detail in Appendix 1. The final 217 phrase dendogram is shown in Appendix 8. A description of the final dendogram, and the aggregation of its branches into a taxonomy of categories, follows. The capitalized phrases in parentheses represent cluster boundary phrases for each category.

The 217 phrases in the dendogram are grouped into 14 clusters. These clusters form the lowest level of the taxonomy hierarchy. Each cluster is assigned a letter, ranging from A to N. The cluster hierarchies are determined by the branch structure. Overall, there are two main branches (clusters). Starting from the phrase adjoining the ‘distance’ ordinate, the first main cluster (A-M) ranges from ELECTROLYTE to ANODE MATERIALS. The second main cluster (N) ranges from ELECTRICAL CONDUCTIVITY to ELECTRONIC CONDUCTIVITY. The clusters differ in size and coherence. The

larger cluster (A-M) covers myriad electrochemical systems and phenomena, while the smaller cluster (N) has a strong focus about materials to improve electrical conductivity and reduce thermal expansion in solid oxide fuel cells.

The smaller cluster (N) will not be sub-divided for the taxonomy, while the larger cluster (A-M) can be divided further into two clusters. The smaller of these two clusters (A-C) ranges from ELECTROLYTE to LITHIATION, while the larger of these two clusters (D-M) ranges from ELECTRODES to ANODE MATERIALS. Again, the clusters differ in size and coherence. The larger cluster (D-M) covers myriad electrochemical systems and phenomena, while the smaller cluster (A-C) has a strong focus about electrolyte and electrode materials to improve electrical conductivity and reversibility of lithium-ion secondary batteries.

The smaller cluster (A-C) can be divided into its elemental clusters. Cluster A (ELECTROLYTE to ELECTROCHEMICAL STABILITY) focuses on the use of polymer electrolytes and carbonate plasticizers to increase ionic conductivity within lithium-ion batteries. Cluster B (LITHIUM to K) focuses on the lithium intercalation into carbon-based electrodes to increase reversible capacity of secondary lithium-ion batteries, with a secondary emphasis on examination of lithium family member intercalation processes. Cluster C (LI IONS to LITHIATION) focuses on electronic and diffusion phenomena related to the intercalation process.

The larger cluster (D-M) can be subdivided further into two clusters, the larger of these two clusters (D-J) ranging from ELECTRODES to NAFION, while the smaller of these two clusters (K-M) ranges from ELECTROCHEMICAL PROPERTIES to ANODE MATERIALS. The larger of these two clusters (D-J) covers myriad electrochemical system types and phenomena, while the smaller of these two clusters (K-M) is focused on improving discharge capacity, reversibility, and cyclability of lithium-ion secondary batteries.

The smaller cluster (K-M) can be sub-divided into its elemental clusters. Cluster K (ELECTROCHEMICAL PROPERTIES to MO) focuses on materials to improve discharge capacity and cycle life of lithium-ion secondary batteries. Cluster L (NICKEL to MOLTEN CARBONATE FUEL CELLS) focuses primarily on materials to improve reversibility of lithium cells, and secondarily on materials to reduce corrosion in molten carbonate fuel cells. Cluster M (ELECTROCHEMICAL PERFORMANCE to ANODE MATERIALS) focuses on material fabrication to improve capacity and cyclability of lithium ion batteries.

The larger cluster (D-J) can be divided further into two clusters, the larger of these two clusters (D-H) ranging from ELECTRODES to ELECTRODEPOSITION, and the smaller of these two clusters (I-J) ranging from CYCLIC VOLTAMMETRY to NAFION. The larger of these two clusters (D-H) covers mainly battery, capacitor, and fuel cell system performance issues, while the smaller of these two clusters (I-J) covers direct methanol oxidation fuel cells and associated thin film characterization issues.

The larger cluster (D-H) can be sub-divided into its elemental clusters. Cluster D (ELECTRODES to ELECTROCHEMICAL CELLS) focuses on characterization of electrode performance in electrochemical cells. Cluster E (DISCHARGE to SIMULATION) focuses on electrode-driven charge-discharge phenomena in rechargeable batteries, including modeling and simulation of battery performance. Cluster F (CAPACITANCE to IMPEDANCE MEASUREMENTS) focuses on preparation of electrode materials by the sol-gel process for use in electrochemical cells in general, and capacitors in particular. Cluster G (CATHODE to SOLID OXIDE FUEL CELLS SOFC) focuses on electrode-related electrochemical cell performance, with emphasis on solid oxide fuel cells. Cluster H (HYDROGEN to ELECTRODEPOSITION) focuses on the direct electrochemical oxidation of methane for solid oxide fuel cells, bypassing the need for hydrogen storage, and addresses carbon deposition reduction at the electrodes by current density variation.

The smaller cluster (I-J) can be sub-divided into its elemental clusters. Cluster I (CYCLIC VOLTAMMETRY to ELECTROCHEMICAL ACTIVITY) focuses on measurement and characterization of thin film electrochemical properties. Cluster J (OXIDATION to NAFION) focuses on methanol oxidation in platinum-ruthenium catalyst–based thin film polymer-electrolyte membrane direct methanol fuel cells.

From these multi-link clustering results, the main research focal points appear to be solid oxide fuel cells (increasing electric/ ionic conductivity, direct hydrocarbon oxidation) and lithium secondary batteries (improving electrode intercalation, increasing electrolyte conductivity, increasing discharge capacity/ reversibility). Research also emphasizes other direct hydrocarbon fuel cells, and other secondary batteries.

4.3.1.3. High Frequency Low Frequency Phrase Relationships

The 217 highest frequency phrases and 7560 lower frequency phrases, were used to form a co-occurrence matrix. The Inclusion Index (Ii=Cij/Ci) was used because the matrix element values remain invariant with the distance from the matrix origin. The resultant associated data served as the basis for finding relationships between low and high frequency phrases. A couple of examples follow, for illustrative purposes.

INTERGRAL RIBS, a low frequency phrase, is strongly related to FUEL CELLS, a high frequency phrase. There is a new design of SOFCs that has a shortened current path, which produces lower cell resistance and higher power output than with tubular cells. This design combines the integral ribs with a flattened air electrode and the seal-less feature of tubular cells.

HOT PHOSPHORIC ACID, a low frequency phrase, is strongly related only to CARBON, a high frequency phrase. The durability of the phosphoric acid fuel cell (PAFC) is strongly affected by the corrosion rate of carbon, an essential material in PAFCs.

While these examples and the associated analysis are very abbreviated due to space limitations, the capability offered is intrinsically very powerful. The high frequency phrase/ low frequency phrase relational technique allows a microcasm of the total text mining study to be performed for any high frequency phrase, or combination of high frequency phrases. Typically, tens, or even hundreds, of low frequency phrases can be identified as relating to a high frequency phrase (depending on the total number of phrases used, and the threshold values of the numerical relational indicators chosen). These low frequency phrases can be categorized into a taxonomy, allowing the technical and bibliometric infrastructure for the high frequency phrase/ concept to be obtained.

4.3.1.4 Recommended Taxonomy

The different statistical and non-statistical taxonomies generated above used different methodologies and some different phrases. Therefore, the results are not directly comparable. A taxonomy that reflects the levels of effort and specific research thrusts would have the structure of the non-statistical Abstract field taxonomy. It would reflect high emphasis on the solid oxide fuel cells and rechargeable lithium batteries, especially the sub-thrusts identified in the statistical clustering approaches. It would also reflect emphasis on other rechargeable battery approaches, such as Ni-M-H, and on direct hydrocarbon fuel cells. Its capacitor component would be low, with emphasis on thin-film voltage-breakdown-resistant super-capacitors.

5. SUMMARY AND DISCUSSION

Two final observations on the technical results. Based on reading a large number of the retrieved SCI Abstracts, and examining the phrase listings and taxonomy results, there appears to be a large imbalance between theory and experiment. The discipline appears almost Edisonian in nature. In addition, there is little evidence of extrapolation of concepts and insights from other technical disciplines. The research appears very parochial. Some of the hybrid literature-based discovery/ multi-discipline workshop techniques (1), performed at the initiation of research projects, could systematically access this extra-disciplinary information.

This paper has presented a number of advantages of using DT and bibliometrics for deriving technical intelligence from the published literature. Large amounts of data can be accessed and analyzed, well beyond what a finite group of expert panels could analyze in a reasonable time period. Preconceived biases tend to be minimized in generating roadmaps. Compared to standard co-word analysis, DT uses full text, not index words, and can make maximum use of the rich semantic relationships among the words. It also has the potential of identifying low occurrence frequency but highly theme related phrases that are 'needles-in-a-haystack', a capability unavailable to any of the other co-occurrence methods.

Combined with bibliometric analyses, DT identifies not only the technical themes and their relationships, but relationships among technical themes and authors, journals, institutions, and countries. Unlike other roadmap development processes, DT generates the roadmap in a 'bottom-up' approach. Unlike other taxonomy development processes, DT can generate many different types of taxonomies (because it uses full text, not key words) in a 'bottom-up' process, not the typical arbitrary 'top-down' taxonomy specification process. Compared to co-citation analysis, DT can use any type of text, not only published literature, and it is a more direct approach to identifying themes and their relationships.

The maximum potential of the DT and bibliometrics combination can be achieved when these two approaches are combined with expert analysis of selected portions of the database. If a manager, for example, wants to identify high quality research thrusts as well as science and technology gaps in specific technical areas, then an initial DT and bibliometrics analysis will provide a contextual view of work in the larger technical area; i.e., a strategic roadmap. With this strategic map in hand, the manager can then commission detailed analysis of selected abstracts to assess the quality of work done as well as identify work that needs to be done (promising opportunities).

5. REFERENCES

1). Kostoff, R. N. (1999). Science and Technology Innovation. Technovation. 19.

2). Kostoff, R. N. (1993). Database Tomography for Technical Intelligence. Competitive Intelligence Review. 4:1.

3). Kostoff, R.N. (1994). Database Tomography: Origins and Applications. Competitive Intelligence Review. Special Issue on Technology, 5:1.

4). Kostoff, R. N. et al (1995). System and Method for Database Tomography. U.S. Patent Number 5440481.

5). Kostoff, R. N., Eberhart, H. J., and Toothman, D. R. (1997). Database Tomography for Information Retrieval. Journal of Information Science. 23:4.

6). Kostoff, R. N. (1997). The Handbook of Research Impact Assessment. Seventh Edition. DTIC Report Number ADA296021. Also, available at http://www.dtic.mil/dtic/kostoff/index.html

7). Kostoff, R. N., Eberhart, H. J., and Toothman, D. R. (1999). Hypersonic and Supersonic Flow Roadmaps Using Bibliometrics and Database Tomography. JASIS. 15 April.

8). Kostoff, R. N., Eberhart, H. J., and Toothman, D. R. (1998). Database Tomography for Technical Intelligence: A Roadmap of the Near-Earth Space Science and Technology Literature. Information Processing and Management. 34:1.

9). Kostoff, R. N., Eberhart, H. J., Toothman, D. R., and Pellenbarg, R. (1997). Database Tomography for Technical Intelligence: Comparative Roadmaps of the Research Impact Assessment Literature and the Journal of the American Chemical Society, Scientometrics, 40:1.

10) Kostoff, R. N., Braun, T., Schubert, A., Toothman, D. R., and Humenik, J. (2000). Fullerene Roadmaps Using Bibliometrics and Database Tomography. Journal of Chemical Information and Computer Science. Jan-Feb.

11) Kostoff, R. N., Green, K. A., Toothman, D. R., and Humenik, J. (2000). Database Tomography Applied to an Aircraft Science and Technology Investment Strategy. Journal of Aircraft, 37:4, July-August.

12). SCI. (1999). Science Citation Index. Institute for Scientific Information. Phila., PA.

13) Kostoff, R. N. (2000). The Underpublishing of Science and Technology Results. The Scientist. 1 May.

14). Garfield, E. (1985) History of Citation Indexes for Chemistry - A Brief Review. JCICS. 25(3). 170-174.

15). Kostoff, R. N. (1998). The Use and Misuse of Citation Analysis in Research Evaluation. Scientometrics, 43:1, September.

16). MacRoberts, M., and MacRoberts, B. (1996). Problems of Citation Analysis. Scientometrics. 36:3. July-August.

6. APPENDICES

APPENDIX 1 – CLUSTERING METHODOLOGIES

Phrase clustering is the grouping of word strings by similarity to some metric. When applied to technical literature (unstructured technical text), it produces groups of technical thrusts. The phrase members of each thrust grouping are related to the group's theme. When applied to a bibliometric list, such as a list of technical paper authors, it produces groups of bibliometric quantities (e.g., authors). In the author case, the authors of each group would have some common similarity, depending on the metric chosen. For example, the groups could be defined by people who publish together, or people who work at the same institution, or people who work in the same topical area.

In the present paper, three phrase clustering approaches are used. One approach is non-statistical, and the other two are statistical. Each approach has the same type of starting point. Text to be analyzed is assembled, either from unstructured free text (e.g., reports, databases with free text fields, etc), or databases with bibliometric fields (e.g., databases of journal papers with author name fields, author institution fields, journal fields, etc). The text to be analyzed is then converted to phrases, with a frequency of occurrence associated with each phrase. The physical locations of the phrases in the source documents are retained, so that this co-occurrence information can be exploited in the grouping process. Now, each of the grouping/ clustering methods will be described in detail.

1A. Non-Statistical Clustering

In this paper, the TextSlicer software (associated with the Database Tomography process and system) was used for the non-statistical clustering phrase frequency analysis. This software generates the frequencies of all single and adjacent double and triple word high technical content phrases contained in databases of interest. It allows multiple counting of nested phrases (i.e., for the double-word phrase METAL MATRIX, METAL is counted as a single word phrase, MATRIX is counted as a single word phrase, and METAL MATRIX is counted as a double word phrase). Through visual inspection and subsequent manual grouping of the technical phrases by technical experts, taxonomies of each database analyzed are then generated, in order to ascertain the specific technical emphases of each database. Two main criteria are employed to select taxonomy categories. First, the categories should be relatively independent and, in aggregate, should allow for rational allocation of all the phrase frequency data. Second, a balance is required between the unwieldiness of too many categories and the insufficient discrimination of too few categories. To quantify the technical emphases, phrases and their associated frequencies of occurrence are binned into the appropriate taxonomy categories based on the expert analyst's judgement. The frequencies of the phrases in each category are summed to provide some estimate of relative levels of global emphasis of each category.

In the present study, and in all studies that have some specific technical focus, two generic types of high technical content phrases result. One type is system or technology specific (e.g., fuel cell, battery, capacitor). The other type is applicable to multiple systems or applications (e.g., thermal expansion, electrical conductivity), and is termed tech base phrases. These tech base phrases can be categorized two ways, depending on the taxonomy objectives. One way is to establish general tech base categories (e.g., properties, phenomena, materials, etc), and manually assign these tech base phrases to the appropriate general categories. The other way is to relate the tech base phrases to the more technology or system specific categories, based on the system-specific context in which they were used in the unstructured text. For the latter assignment of tech base phrases into system-specific categories, phrase proximity analysis is required.

For a system or technology-specific technical phrase (e.g., fuel cell), phrase proximity analysis generates a dictionary of phrases located in close physical proximity to the system-specific phrase throughout the text. Each phrase in the dictionary has a number of associated numerical indicators, including its frequency of occurrence throughout the text. These dictionary phrases are then assigned to the system-specific category in the taxonomy. Depending on the objectives of the study, tech base phrase duplication in categories may be allowed, or adjustments may have to be made across all categories to eliminate phrase duplication (multiple counting). In the present paper, the phrase proximity analysis algorithm of the TextSlicer software was used.

1B. Statistical Clustering

The statistical clustering approaches start with the phrase frequency and location information, then group phrases based on their co-occurrence frequencies and other important numerical indicators. In this paper, the statistical clustering has used two software packages in tandem. The phrase frequency and location information is generated by TechOasis, a software package from Search Technology. TechOasis uses Natural Language Processing (NLP) to extract the phrases and their frequencies from the free text. Not all phrases are extracted (typical of all NLP), and manual cleanup of the extracted phrases is still required to eliminate the lower technical content phrases. Also, unlike the multiple counting of nested phrases in TextSlicer, multiple counting of nested phrases in TechOasis is disallowed (i.e., METAL MATRIX would be counted once as a double word phrase, but neither METAL nor MATRIX would be counted as single word phrases). The resulting phrase frequency and co-occurrence matrix information is then exported from TechOasis to WINSTAT, an Excel add-in software package. Two types of output are generated by WINSTAT: factor matrix and multi-link clustering. Each of these processes will now be described.

1B-1. Multi-Link Clustering

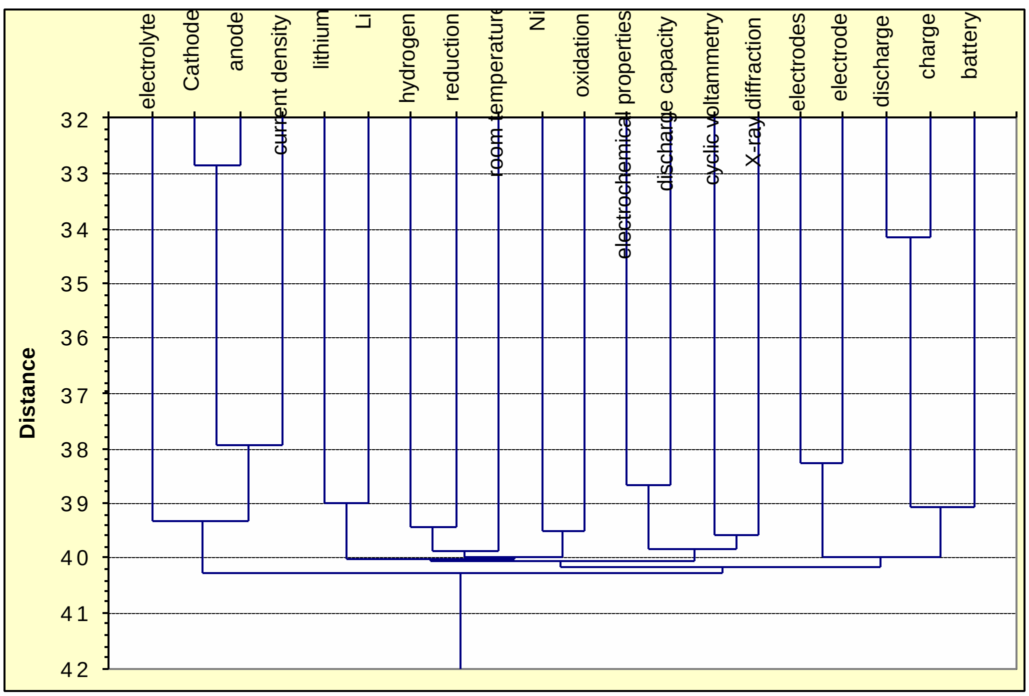
Phrase multi-link clustering starts with pre-defined metrics (variables), and groups phrases (cases) based on the strength of their relationships with these metrics. For example, if a symmetric phrase-phrase matrix is used as the basis for clustering, then

the closeness of two phrases (cases) will be determined by their co-occurrence profile with all other phrases (variables). The clustering becomes a co-occurrence profile matching process. The WINSTAT software package produces three types of related outputs.

1) A dendogram that shows the quantitative linkages among closely-related phrases. Figure A1B-1, for example, is a dendogram that portrays linkages among the twenty highest frequency technical content phrases from the Abstracts database.

   A dendogram is a tree-like structure that shows linkages between phrases. It does so by starting with a root that encompasses all the phrases (See the vertical line on Figure A1B-1 ranging from a Distance value of 42 to slightly over 40). Then it splits into two groups (clusters) until all the phrases are contained in their own cluster. In Figure A1B-1, the root at the bottom of the page encompasses all the phrases. The first split is into two large clusters. One cluster contains the phrases ELECTROLYTE, CELL, CATHODE, ANODE, and CELLS. The second cluster contains all the remaining phrases LITHIUM, ELECTROCHEMICAL PROPERTIES, CYCLIC VOLTAMMETRY, X-RAY DIFFRACTION, ELECTRODES, ELECTRODE, HYDROGEN, ALLOY, ALLOYS, BATTERY, AIR, OXYGEN, OXIDATION, WATER, and CONDUCTIVITY.

FIGURE A1B-1 – EXAMPLE TWENTY PHRASE DENDOGRAM



2) A table that contains a quantitative measure of the similarity of adjoining phrases or phrase-cluster pairs. The similarity, or 'distance', is obtained by matching the profiles as described above. Figure A1B-2, for example, is a table that contains the information portrayed in Figure A1B-1. The distances shown on the dendogram are taken from the distances given in this table; thus, the table is the numerical expression of the dendogram.

FIGURE A1B-2 – ELEMENTAL STEPS IN DENDOGRAM FORMATION

joining Cluster 1	**Size 1**	**with Cluster 2**	**Size 2**	**Distance**
Cathode	1	anode	1	32.83216927
discharge	1	charge	1	34.16843369
Cathode	2	current density	1	37.9232786
electrodes	1	electrode	1	38.29297045
electrochemical properties	1	discharge capacity	1	38.68748862
lithium	1	Li	1	39.00662295
discharge	2	battery	1	39.06016967
electrolyte	1	Cathode	3	39.33157288
hydrogen	1	reduction	1	39.42153664
Ni	1	oxidation	1	39.52735747
cyclic voltammetry	1	X-ray diffraction	1	39.58094847
electrochemical properties	2	cyclic voltammetry	2	39.85159646
hydrogen	2	room temperature	1	39.86541946
hydrogen	3	Ni	2	39.97396716
electrodes	2	discharge	3	39.97805457
lithium	2	hydrogen	5	40.00961352
lithium	7	electrochemical properties	4	40.04779843
lithium	11	electrodes	5	40.17291601
electrolyte	4	lithium	16	40.26135359

3) A taxonomy of a pre-specified number of groups of phrases. Figure A1B-3, for example, shows the groupings of phrases when four clusters were specified for the data portrayed in Figure A1B-1.

FIGURE A1B-3

Cluster #	Phrases
1	electrolyte
2	lithium
3	electrodes
3	electrode
2	Li
1	Cathode
2	hydrogen
2	room temperature
4	electrochemical properties
4	cyclic voltammetry
2	reduction
1	anode
3	discharge
4	X-ray diffraction
2	Ni
4	discharge capacity
3	battery
3	charge
2	oxidation
1	current density

Final categories can be generated two ways. The clusters can be manually grouped into categories. Or, the number of categories (clusters) can be specified to the algorithm, and the computer output will contain the final categories. Usually, some modest manual re-grouping of the computer output groups is required to arrive at a final recommended taxonomy.

1C. FACTOR MATRIX CLUSTERING

Factor matrix clustering generates a correlation matrix from the frequency and location information of the phrases. It then generates factors that are composed of all the phrases in the correlation matrix. The phrases are ordered quantitatively by their correlation with each other, with the most strongly correlated assigned the highest quantitative values.

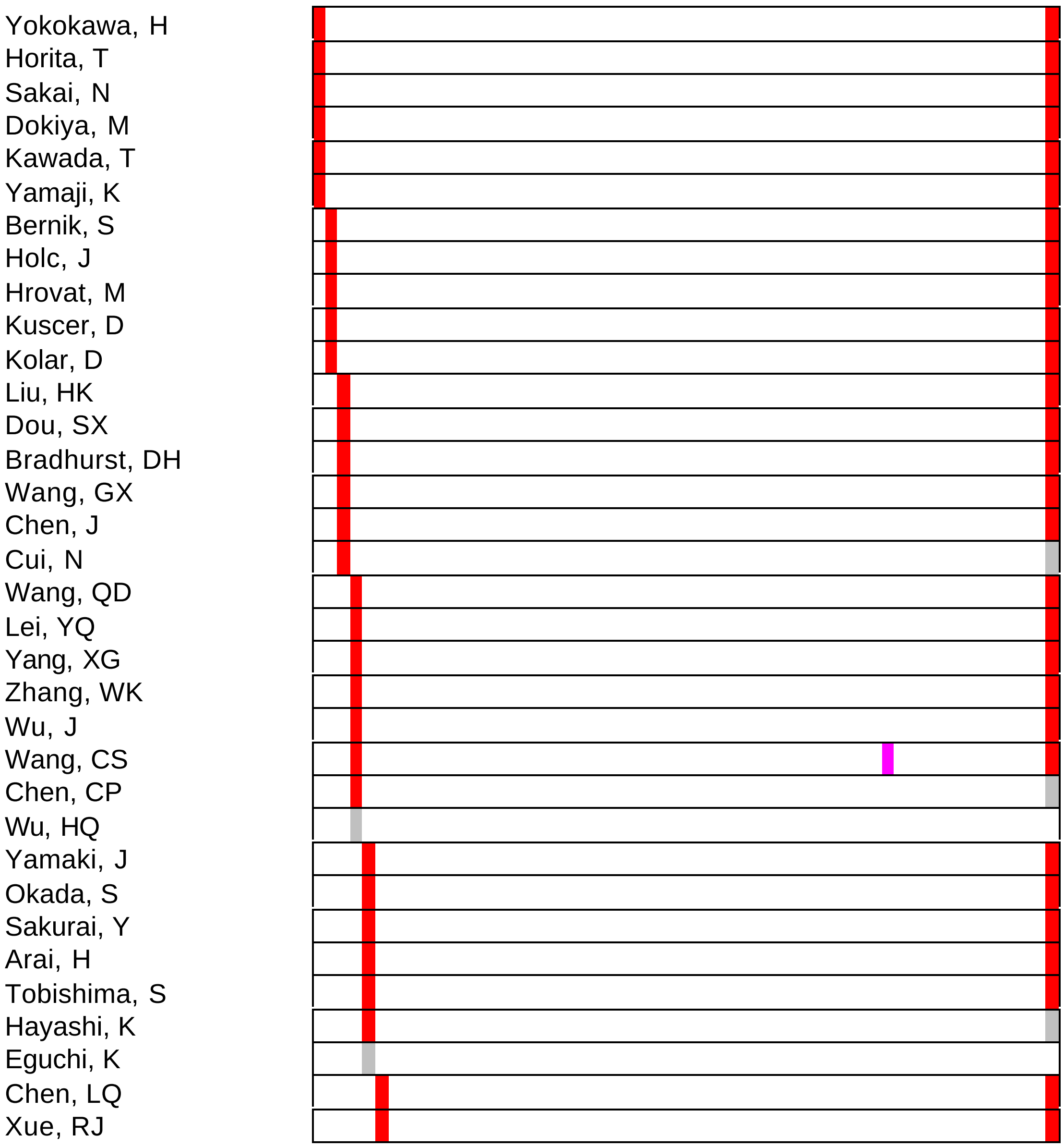
Appendix 2 shows an author factor matrix. The grouping metric was the number of co-authored publications, taking into account the total number of publications of each author. In order to display all factors on one page width, the column widths had to be shrunk drastically. The matrix element values were hidden, but the shadings could be displayed, and they represent high correlation values (factor loadings). The darker the shading, the higher the value.

The number of factors generated can be fully discretionary (e.g., select a ten factor matrix, with no floor on the eigenvalues), or partially discretionary (e.g., set an eigenvalue floor of unity, let the algorithm generate the final number of factors). Final

taxonomies can be generated by manually combining factors into categories, or specifying the number of categories (factors) desired to the algorithm.

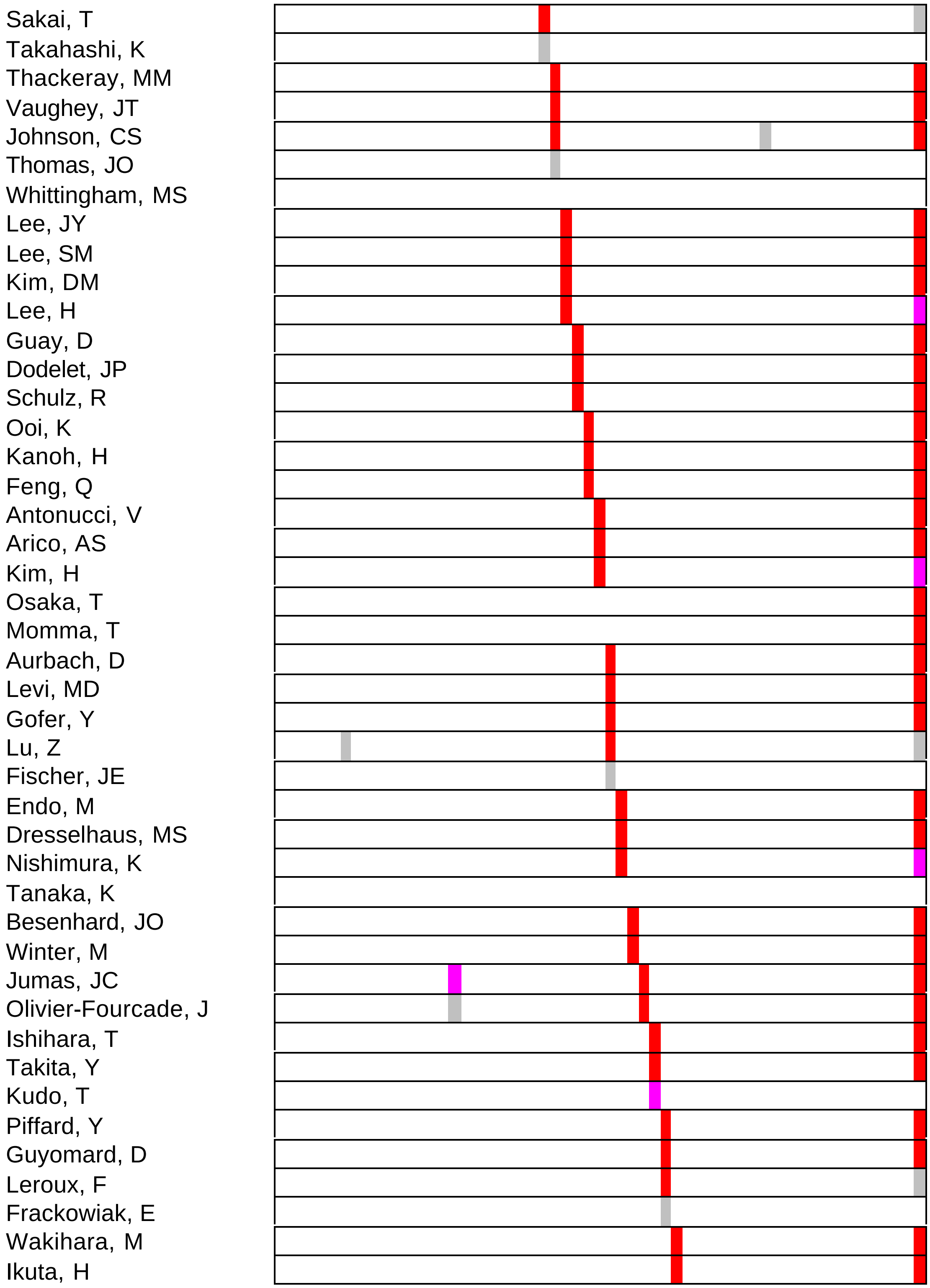
The author factor matrix in Appendix 2 was generated by selecting an eigenvalue floor of unity. In practice, this means that each factor generated will add new information. Analysis of the author factor matrix in Appendix 2 is explained in the author bibliometrics section of the text, and will not be repeated here.

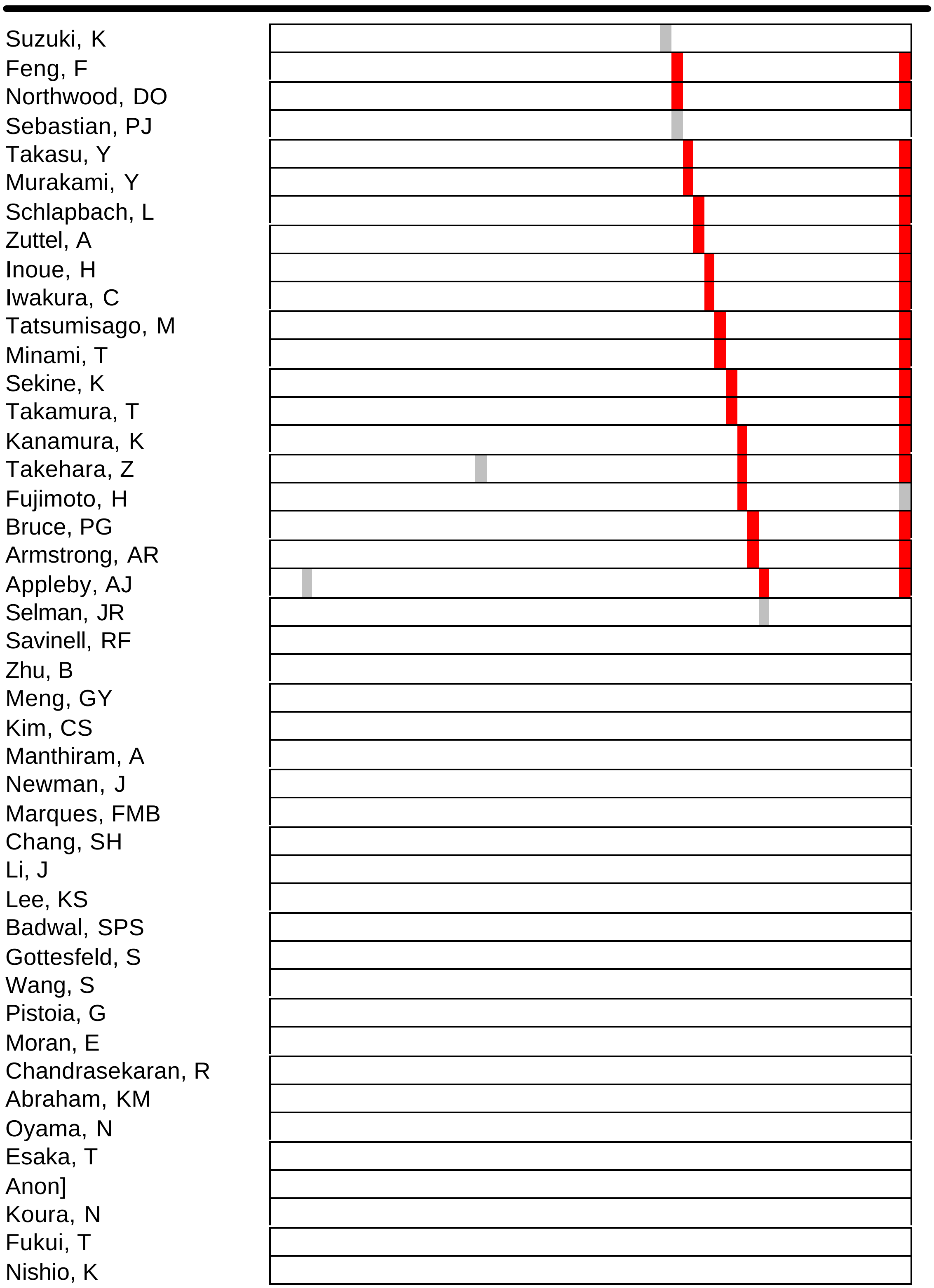
APPENDIX 2 – AUTHOR FACTOR MATRIX

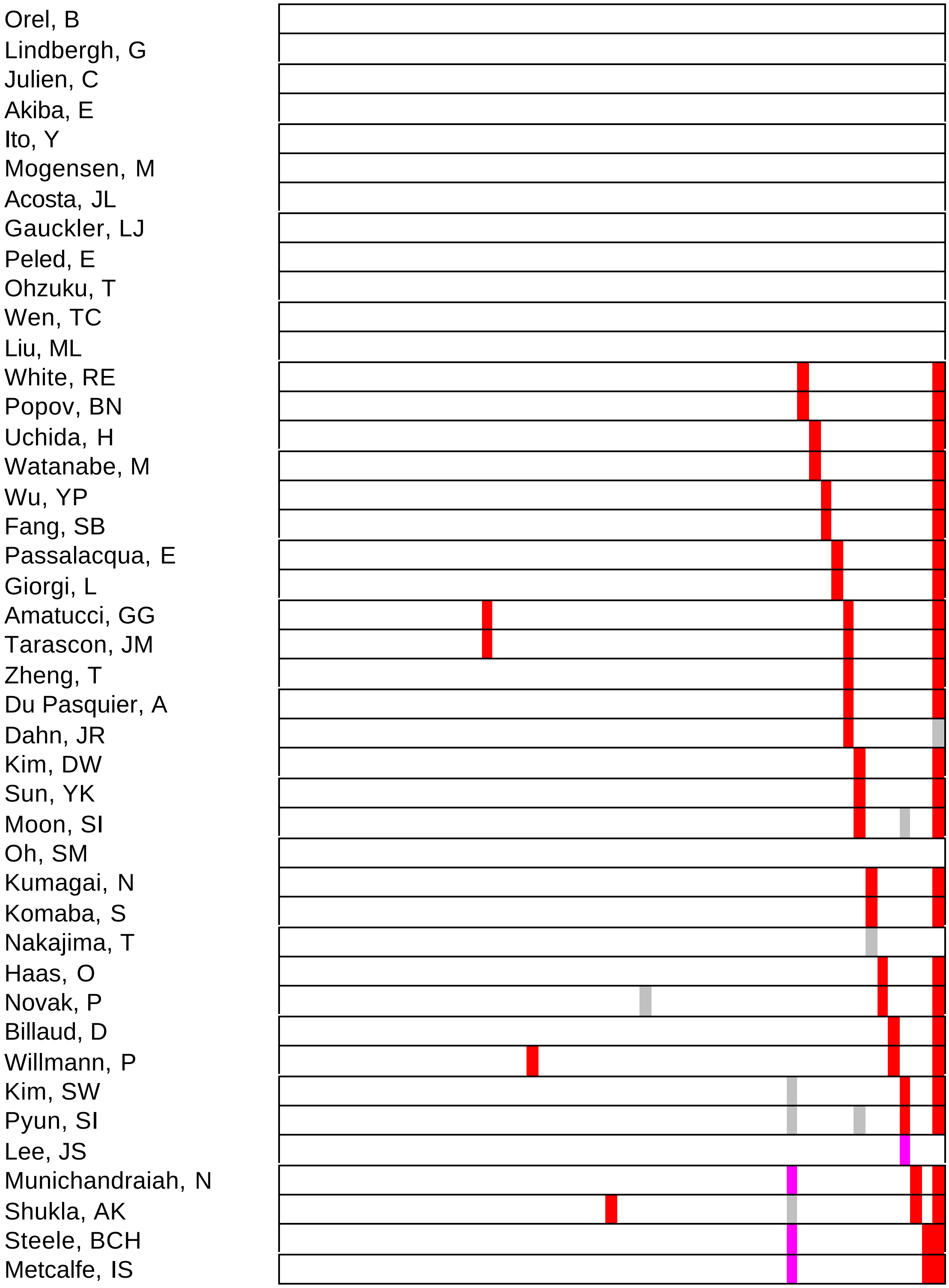



Wang, ZX
Huang, XJ
Huang, H
Song, DY
Zhang, YS
Gao, XP
Yuan, HT
Liu, J
Amine, K
Kageyama, H
Tabuchi, M
Kobayashi, H
Kanno, R
West, AR
Goodenough, JB
Jang, YI
Sadoway, DR
Chiang, YM
Huang, BY
Ceder, G
Yamamoto, O
Takeda, Y
Imanishi, N
Yang, J
Mori, M
Sammes, NM
Watanabe, T
Baffier, N
Bach, S
Pereira-Ramos, JP
Farcy, J
PereiraRamos, JP
Scrosati, B
Appetecchi, GB
Croce, F
Panero, S
Mastragostino, M
Koksbang, R
Barker, J
Saidi, MY
West, K
Kelder, EM
Schoonman, J
Jak, MJG

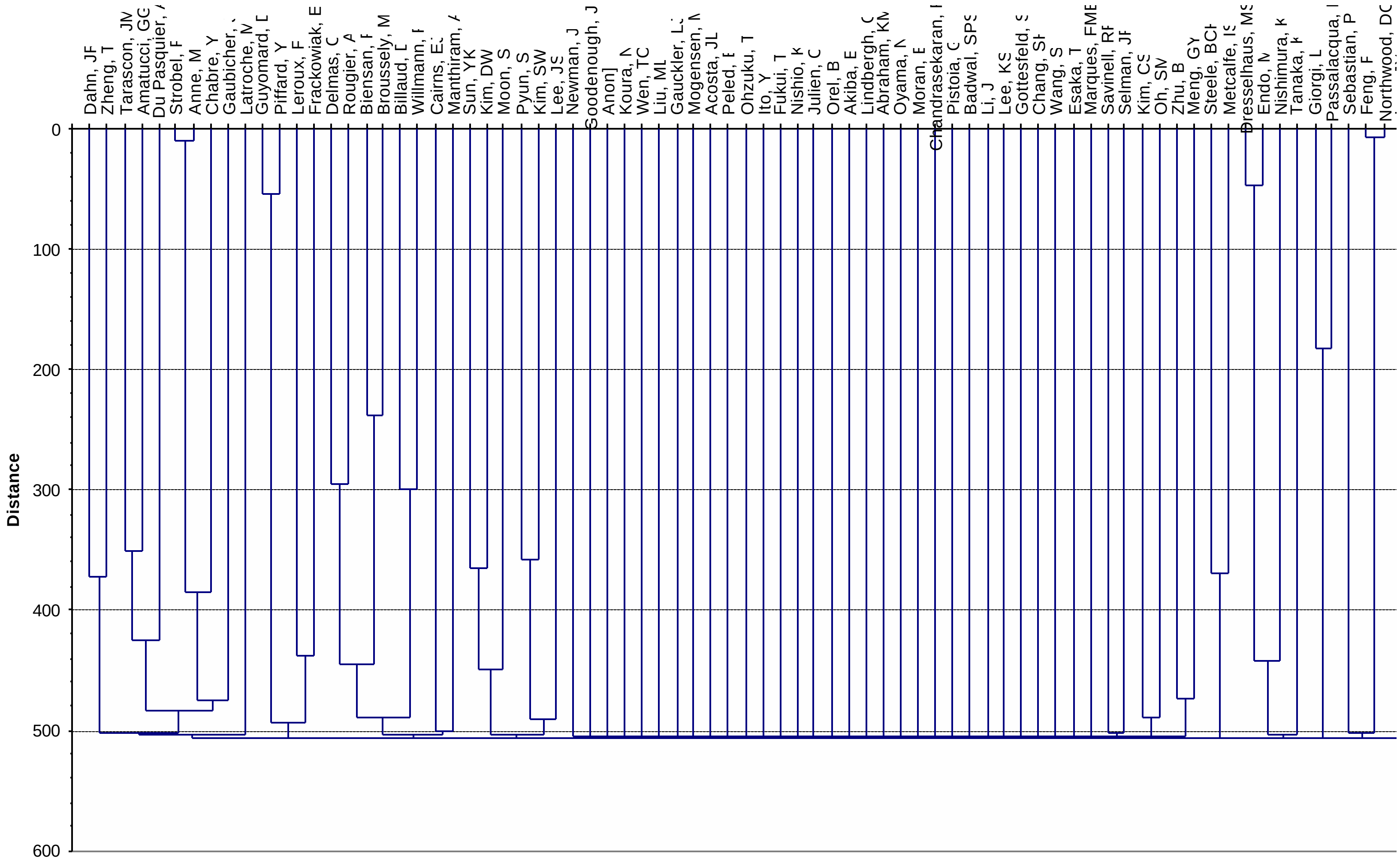

Hemmes, K
Taama, WM
Argyropoulos, P
Scott, K
Miura, T
Kishi, T
Kawakita, J
Morales, J
Tirado, JL
Sanchez, L
Lavela, P
McBreen, J
Mukerjee, S
Yang, XQ
Srinivasan, S
Ticianelli, EA
Ross, PN
Anne, M
Strobel, P
Chabre, Y
Gaubicher, J
Latroche, M
Inaba, M
Ogumi, Z
Abe, T
Uchida, I
Itoh, T
Nishizawa, M
Yamada, K
Passerini, S
Smyrl, WH
Owens, BB
Prosini, PP
Broussely, M
Biensan, P
Rougier, A
Delmas, C
Cairns, EJ
Morita, M
Ishikawa, M
Matsuda, Y
Yoshio, M
Xia, YY
Noguchi, H


Sakai, T
Takahashi, K
Thackeray, MM
Vaughey, JT
Johnson, CS
Thomas, JO
Whittingham, MS
Lee, JY
Lee, SM
Kim, DM
Lee, H
Guay, D
Dodelet, JP
Schulz, R
Ooi, K
Kanoh, H
Feng, Q
Antonucci, V
Arico, AS
Kim, H
Osaka, T
Momma, T
Aurbach, D
Levi, MD
Gofer, Y
Lu, Z
Fischer, JE
Endo, M
Dresselhaus, MS
Nishimura, K
Tanaka, K
Besenhard, JO
Winter, M
Jumas, JC
Olivier-Fourcade, J
Ishihara, T
Takita, Y
Kudo, T
Piffard, Y
Guyomard, D
Leroux, F
Frackowiak, E
Wakihara, M
Ikuta, H

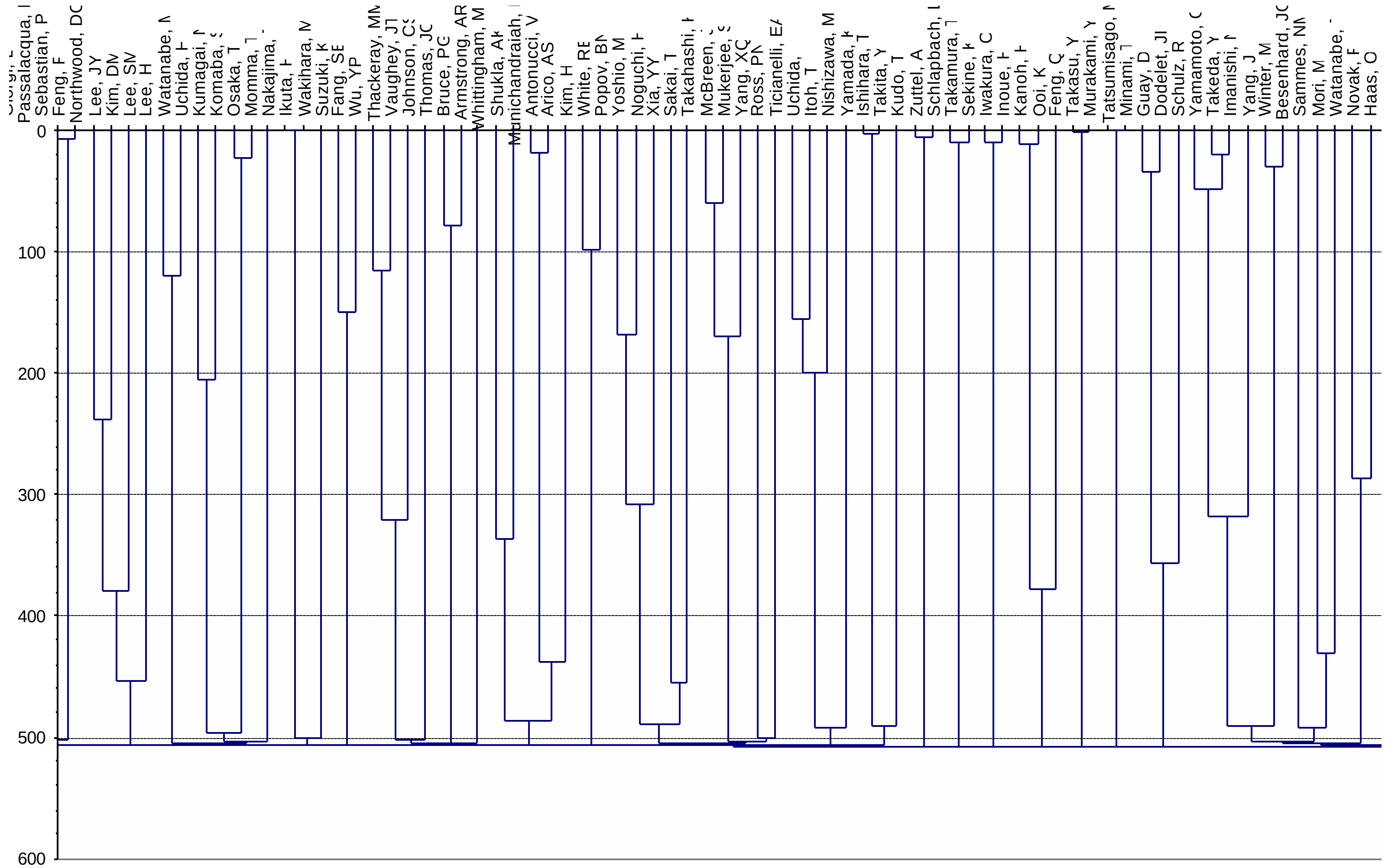

Suzuki, K
Feng, F
Northwood, DO
Sebastian, PJ
Takasu, Y
Murakami, Y
Schlapbach, L
Zuttel, A
Inoue, H
Iwakura, C
Tatsumisago, M
Minami, T
Sekine, K
Takamura, T
Kanamura, K
Takehara, Z
Fujimoto, H
Bruce, PG
Armstrong, AR
Appleby, AJ
Selman, JR
Savinell, RF
Zhu, B
Meng, GY
Kim, CS
Manthiram, A
Newman, J
Marques, FMB
Chang, SH
Li, J
Lee, KS
Badwal, SPS
Gottesfeld, S
Wang, S
Pistoia, G
Moran, E
Chandrasekaran, R
Abraham, KM
Oyama, N
Esaka, T
Anon]
Koura, N
Fukui, T
Nishio, K


Orel, B
Lindbergh, G
Julien, C
Akiba, E
Ito, Y
Mogensen, M
Acosta, JL
Gauckler, LJ
Peled, E
Ohzuku, T
Wen, TC
Liu, ML
White, RE
Popov, BN
Uchida, H
Watanabe, M
Wu, YP
Fang, SB
Passalacqua, E
Giorgi, L
Amatucci, GG
Tarascon, JM
Zheng, T
Du Pasquier, A
Dahn, JR
Kim, DW
Sun, YK
Moon, SI
Oh, SM
Kumagai, N
Komaba, S
Nakajima, T
Haas, O
Novak, P
Billaud, D
Willmann, P
Kim, SW
Pyun, SI
Lee, JS
Munichandraiah, N
Shukla, AK
Steele, BCH
Metcalfe, IS

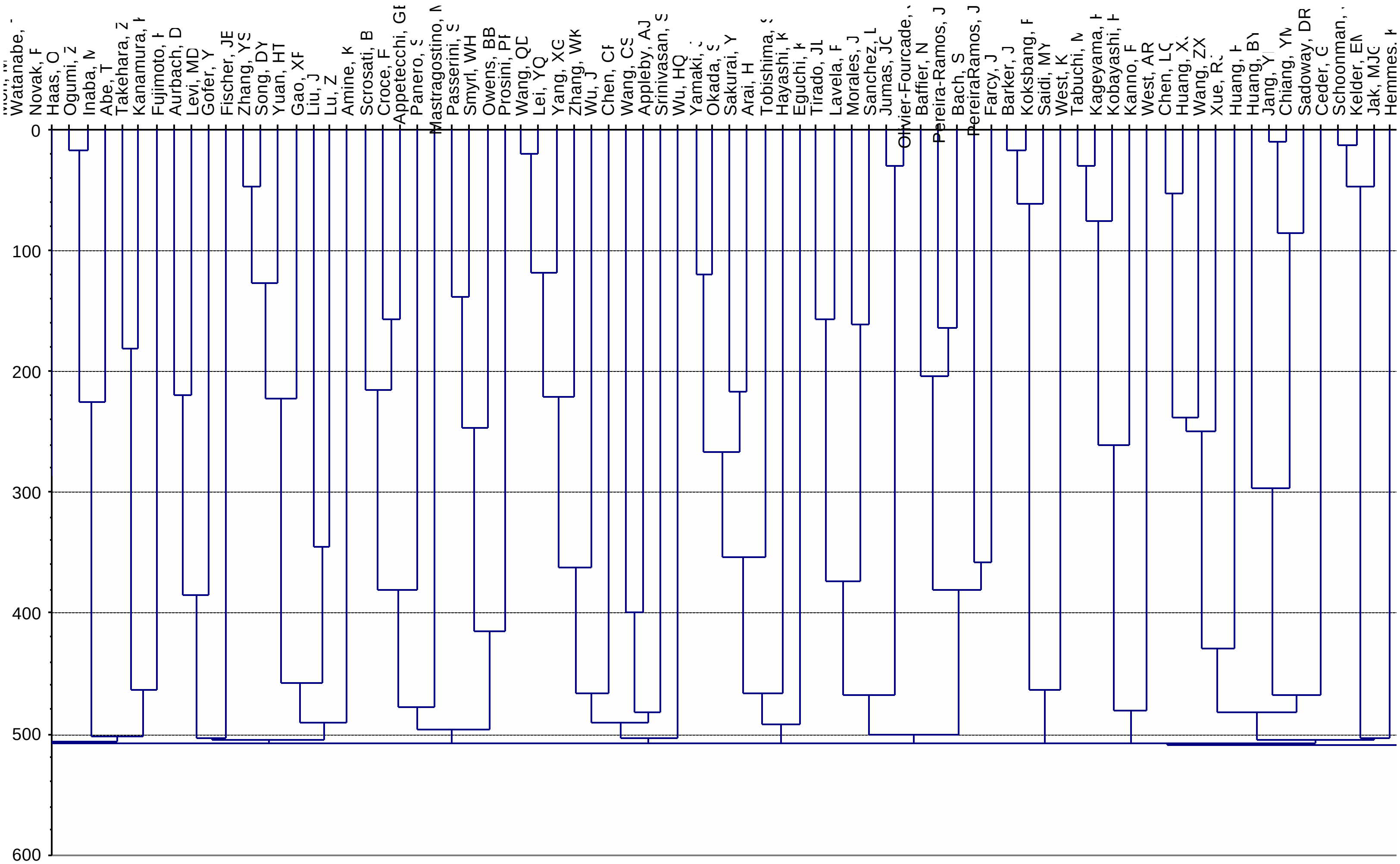
APPENDIX 3 – AUTHOR CLUSTERING DENDOGRAM



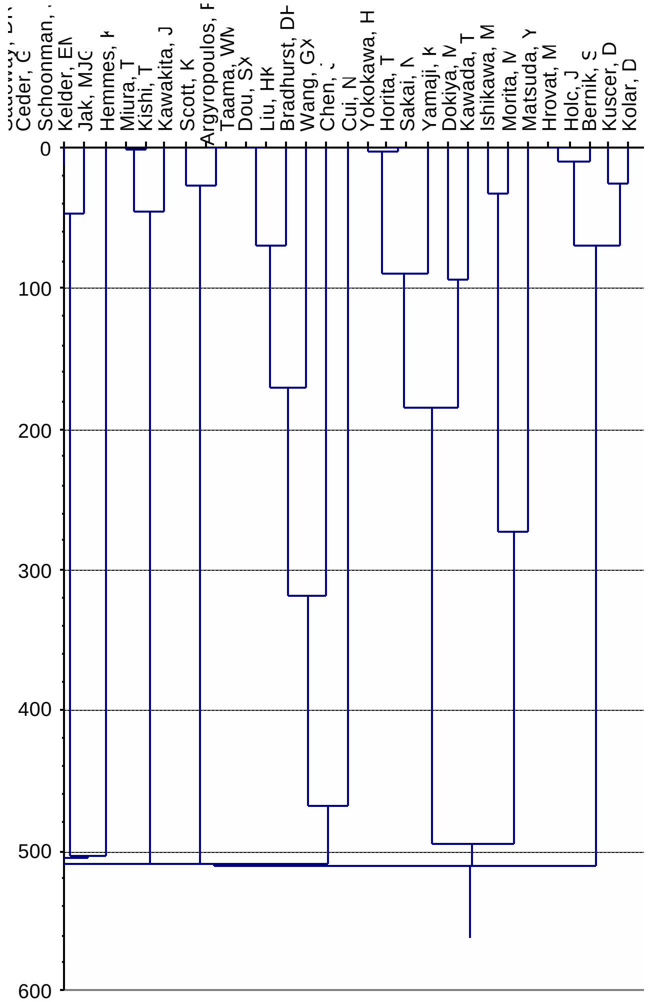
APPENDIX 3 – AUTHOR CLUSTERING DENDOGRAM (CONTINUED)



APPENDIX 3 – AUTHOR CLUSTERING DENDOGRAM (CONTINUED)



APPENDIX 3 – AUTHOR CLUSTERING DENDOGRAM (CONTINUED)



APPENDIX 4 – COUNTRY CO-PUBLISHING MATRIX

	# Records	1552	1318	558	499	380	344	341	318	250	249	206	151	126	121	113
# Records	**COUNTRY**	**JAPAN**	**USA**	**FRANCE**	**CHINA**	**SOUTH KOREA**	**UK**	**GERMANY**	**CANADA**	**ITALY**	**INDIA**	**RUSSIA**	**SPAIN**	**SWEDEN**	**AUSTRALIA**	**SWITZERLAND**
1552	**JAPAN**	1552	52	15	17	16	14	8	3	5	4	0	0	3	3	5
1318	**USA**	52	1318	36	6	17	10	9	24	27	9	6	5	5	2	4
558	**FRANCE**	15	36	558	3	4	10	13	9	9	5	4	31	3	0	4
499	**CHINA**	17	6	3	499	1	2	5	5	1	0	0	1	8	0	0
380	**SOUTH KOREA**	16	17	4	1	380	0	4	1	5	2	1	0	0	2	1
344	**UK**	14	10	10	2	0	344	9	7	2	5	2	0	4	2	3
341	**GERMANY**	8	9	13	5	4	9	341	3	7	5	6	0	2	0	13
318	**CANADA**	3	24	9	5	1	7	3	318	0	0	1	0	2	1	1
250	**ITALY**	5	27	9	1	5	2	7	0	250	5	2	1	2	0	2
249	**INDIA**	4	9	5	0	2	5	5	0	5	249	0	1	0	4	2
206	**RUSSIA**	0	6	4	0	1	2	6	1	2	0	206	2	2	0	1
151	**SPAIN**	0	5	31	1	0	0	0	0	1	1	2	151	0	0	0
126	**SWEDEN**	3	5	3	8	0	4	2	2	2	0	2	0	126	0	1
121	**AUSTRALIA**	3	2	0	0	2	2	0	1	0	4	0	0	0	121	0
113	**SWITZERLAND**	5	4	4	0	1	3	13	1	2	2	1	0	1	0	113

APPENDIX 5 – KEYWORDS NON-STATISTICAL TAXONOMY

The main system categories were Sources-Converters, Storage, and Other Applications. Sources-Converters included Fuel Cells, Storage included Batteries and Capacitors (to a much smaller extent), and Other Applications covered many uses of Electrochemical Cells.

Fuel Cells
Solid oxide, molten carbonate, polymer electrolyte, and direct methanol, with modest emphasis on alkaline, proton exchange membrane, and phosphoric acid.

Batteries
Lithium (rechargeable, ion, polymer, polypyrrole, composite, iron sulfide, thionyl chloride), lead-acid (valve-regulated), secondary, rechargeable, nickel-metal-hydride, with modest emphasis on ion shuttlecock, alkaline, nickel hydrogen, aluminum air, polymer, solid-state, and rocking chair.

Capacitors
Electrochemical, supercapacitors, and double-layer, with modest emphasis on ferroelectric, pseudocapacitors, electrolytic (especially aluminum), hybrid, high temperature, thin film, ultracapacitors, and solid state.

Other Applications
Electrochemical cells (especially lithium, lithium-ion, insertion, galvanic, with modest emphasis on hydrogen concentration, grid corrosion balancing, li-socl2, electrolyte, oxygen concentration, thin-layer, non-aqueous, ion-insertion, v2o5-p2o5, nickel-oxide, metal hydride air, cylindrical, chlor-alkali, mg/ mo2 primary, al-cl2, alkaline manganese, na/ sec14, cadmium, linixmn2-xo4/li electrochemical, li-so2, li-ag chromate cells and pacemaker devices, dioxide, manganese dioxide, lithium-carbon, nickel/ aluminum, ni/ fe, nickel-hydrogen, thermogalvanic, rechargeable zn/znso(4)/mno2, solid-state mg/mno2, and sodium/ nickel chloride), and electric vehicles, aircraft, charge storage, memory, medical, military portable equipment, submarines, hydrogen storage, photovoltaic, and space.

Tech-Base Related
The main Tech-Base categories were Electrolytes, Electrodes, Materials, Processes, Properties, Phenomena, Parameters, Experiments, Theory, Macrostructure, Microstructure, and Region.

Electrolytes
Polymer, solid, gel, organic, composite, nonaqueous, 1-i, ceramic, propylene carbonate, ceo2, lanthium gallate, zirconia-based, rubbery, with modest emphasis on composite film, acid, polyurethane-based, ysz-cgo bilayer, carbonate, alkaline, aqueous, inorganic, molten salt, lithium, beta alumina, glass polymer composite, lagao3, polyether, y2o3-stabilized zro2, perovskite solid, ceria, mg2+-conducting gel polymer, alkaline solid polymer, nanocomposite polymer, amorphous, nacl, (ceo2)(1-x)(gdo1.5)(x), liclo4-

ethylene carbonate, h2so4, fluoroethylene carbonate, nitrate, peo-pc-liclo4, and vanadium sulfuric acid.

Electrodes

Metal hydride, gas diffusion, platinum, oxide, lithium, air, porous, rotating disk, composite, oxygen, carbon, hydroxide, noble metal, graphite, insertion, nickel, alloy, polymer-coated, ruo2, micro, redox polymerization, film, counter-, au, pbo2-pbso4, pt/lambda-mno2, nickel hydroxide, oxygen, intercalation, lithium alloy composite, lani5, lead dioxide, zinc, with modest emphasis on hydrogen storage alloy, bpg, au/ na-c, pt/ c, pt/ sn, pattern, rough capacitive, zdf, organic, membrane, powder, transparent, ultra-micro-, highly dispersed, linio2 positive, inert, multinary alloy, hydrogen, cl-2, carbon fiber, glassy carbon, manganese dioxide, ion-selective, spinel, zeolite-modified, polymerization, vanadium oxide, single crystal, nickel hydroxide, ruthenium, titanium, pbo2, meo2hx(yh2o), cation-selective, cermet, cro2, corannulene, cloth polarizable, nafion-coated, anatase, aerogel, au-nafion, molybdenum nitride, lithium cobalt oxide, liquid membrane, iridium dioxide, fullerene film, protoporphyrin modified, polythiophene-modified, and valinomycin.

Materials

-Elements (especially lithium, carbon, platinum, graphite, hydrogen, nickel, oxygen, tin, magnesium, cobalt, ruthenium, sodium, iron, lead, copper, lanthanium, manganese, sulfur, vanadium);

-Compounds (especially oxides, limn2o4, licoo2, linio2, propylene carbonate, water, methanol, polyaniline, methane, polyethelyne oxide, manganese dioxide, zirconia, polypyrrole, manganese oxide, ceria, carbon monoxide, nafion, v2o5, carbonate, hydride, lani5, ruo2, mg2ni, formic acid, lanthium gallate, molten carbonate, alkaline-solution, phosphoric acid, la1-xsrxmno3, ethelyne carbonate, metal hydride, lithium perchlorate, zirconium oxide, limno2, sulfuric acid);

-Classes (especially spinel, alloys, perovskite, complexes, acid, polymers, conducting polymers, composites, particles, metal, ceramics, powders, additives, salts, carbonaceous materials, disordered carbons);

-Functional (especially hydrogen storage alloy, cathode materials, catalysts, electrocatalyst, intercalation compounds).

Processes

Intercalation (especially lithium), insertion (especially lithium), electrochemistry, electrocatalysis, discharge, storage (especially hydrogen), mechanical alloying, deposition (especially electro-), electrochemical characterization, electrochemical intercalation, deintercalation, surface modification, surface treatment, and sintering.

Properties

Conductivity (especially ionic and electrical), electrochemical properties, impedance, capacity, stability, electrical-properties, cycling efficiency, impedance (especially electrochemical and ac), charge, energy, cycle life, safety, discharge capacity, and energy density.

Phenomena

Kinetics, oxidation (especially electro-, methanol, electrocatalytic, electrochemical), oxygen reduction, diffusion, transport, mechanism, adsorption, diffraction, absorption (especially hydrogen), nonstoichiometry, corrosion, growth, dissolution, transition, activation, thermal expansion, solubility, evolution (especially oxygen, hydrogen), polarization, reactivity, exchange, passivation, decomposition, and degradation.

Parameters

Room/ low/ high temperature, and high/ partial pressure.

Experiments

Spectroscopy (especially impedance/ electrochemical impedance, raman), x-ray diffraction, cyclic voltammetry, xps, sol-gel method, and nmr.

Theory

Very modest compared to experiment, and focused on model (especially mathematical) and simulation.

Macrostructure

Films (especially thin), membranes, crystal structure, defect structure, and membrane.

Microstructure

Ion, electronic structure, ad-atoms (especially ruthenium), micro- and nano- scale quantities (nanoparticles, nanocrystalline, carbon nanotubes, nanocomposites, nanomaterials, nanostructure, microbeads, microelectrodes).

Region

Interface, surface, and boundary (especially grain and three phase).

APPENDIX 6 – ABSTRACT MANUAL TAXONOMY

TABLE A6-1 - ABSTRACT TAXONOMY

LEVEL 1	LEVEL 2	LEVEL 3	LEVEL 3 (DETAILED EXAMPLES)
ECHEM CONVERTER (17,227)	FUEL CELLS (17,227)	Fuel cell components and component configurations (4,038)	ELECTRODES (POROUS ELECTRODES, OXYGEN ELECTRODE, MEMBRANE ELECTRODE, ROTATING DISK ELECTRODE), CATHODES, ANODES (CERMET ANODE), MEMBRANES , CATALYSTS (CATHODE CATALYST, ELECTROCATALYSTS), ELECTROLYTES (POLYMER ELECTROLYTE, YSZ ELECTROLYTE), COLLECTORS, DIFFUSION LAYERS, BIPOLAR PLATES, STEAM REFORMER, FUEL PROCESSOR, ELECTROLYZERS, CATALYST LAYERS, STACKS (CELL STACK), PLANAR, RECTANGULAR
		Fuel cell properties and characteristics (2,682)	CONDUCTIVITY(ION CONDUCTIVITY), CURRENT DENSITY, RESISTANCE, POWER DENSITY, VOLTAGE, MASS, IMPEDANCE, PERMEABILITY, FLOW RATES, CORROSION RATE, ELECTROCHEMICAL, POROUS, IONIC, CATALYTIC, STABILITY, AREA, SMALL, STABILIZED, EFFICIENCY, DOPED, THIN, TUBULAR, SINTERED, OHMIC, BIPOLAR, ,ROTATING, PROTON CONDUCTING
		Fuel cell energy sources / fuels (2,385)	GAS, HYDROGEN, METHANOL, OXYGEN, METHANE, NATURAL GAS, ETHANOL, HYDROCARBONS, GASOLINE
		Fuel cell component materials (2,358)	POLYMER, NAFION, YSZ, ZIRCONIA. PLATINUM. CARBON, NICKEL, CERAMICS, CERMETS, CERIA, COMPOSITE, YTTRIA-STABILIZED ZIRCONIA, RU, LANTHANUM, STEEL, MANGANITE, SIO2, CHROMITE

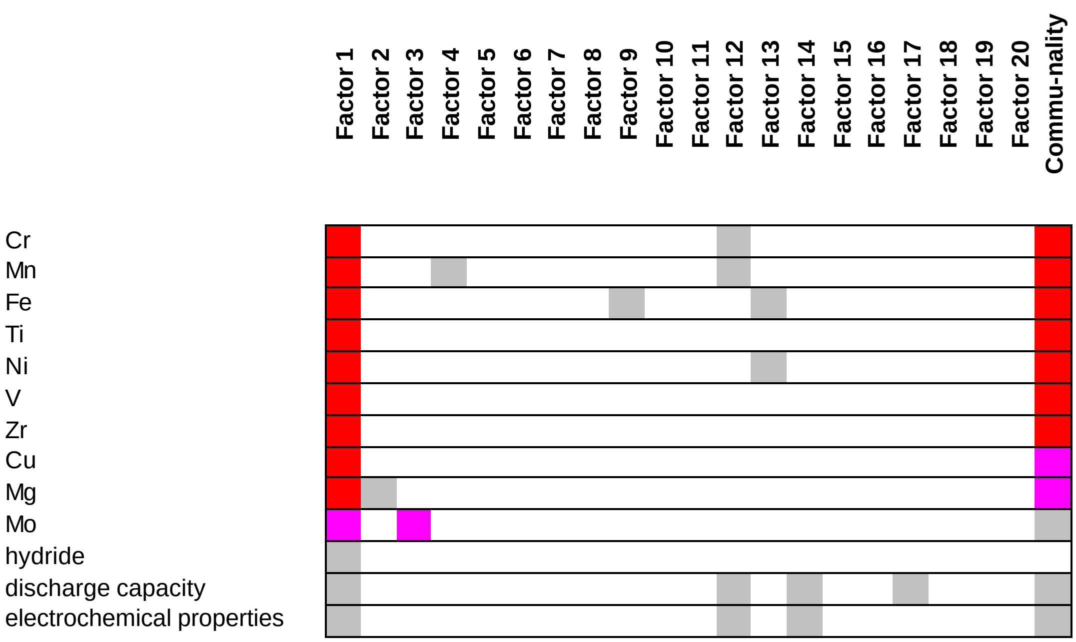
		Fuel cell types (2,238)	SOLID OXIDE (SOFC), MOLTEN CARBONATE (MCFC, DIR-MCFC), DIRECT METHANOL (DMFC, DMFCS), PROTON EXCHANGE MEMBRANE (PEM, PEMFC), POLYMER ELECTROLYTE FUEL (PEFC), POLYMER ELECTROLYTE MEMBRANE, PHOSPHORIC ACID, SOLID POLYMER, ALKALINE
		Fuel cell conversion processes (1,363)	OXIDATION (METHANOL OXIDATION, ELECTROOXIDATION, PARTIAL OXIDATION, DIRECT OXIDATION), REFORMING (STEAM REFORMING, INTERNAL REFORMING), POLARIZATION, OXYGEN REDUCTION, GAS DIFFUSION, SINTERING, CORROSION, THERMAL EXPANSION, HUMIDIFICATION, ELECTROLYSIS, GAS FLOW, REDUCTION REACTION, HYDROGEN PRODUCTION
		Fuel cell conversion process byproduct (1, 011)	WATER, CURRENT, HEAT, CO2 (CARBON DIOXIDE), CARBON MONOXIDE
		Fuel cell operating conditions (885)	TEMPERATURES (OPERATING TEMPERATURE), PRESSURE (PARTIAL PRESSURE)
		Fuel cell application (267)	ELECTRICITY, VEHICLES, POWER GENERATION, TRANSPORTATION

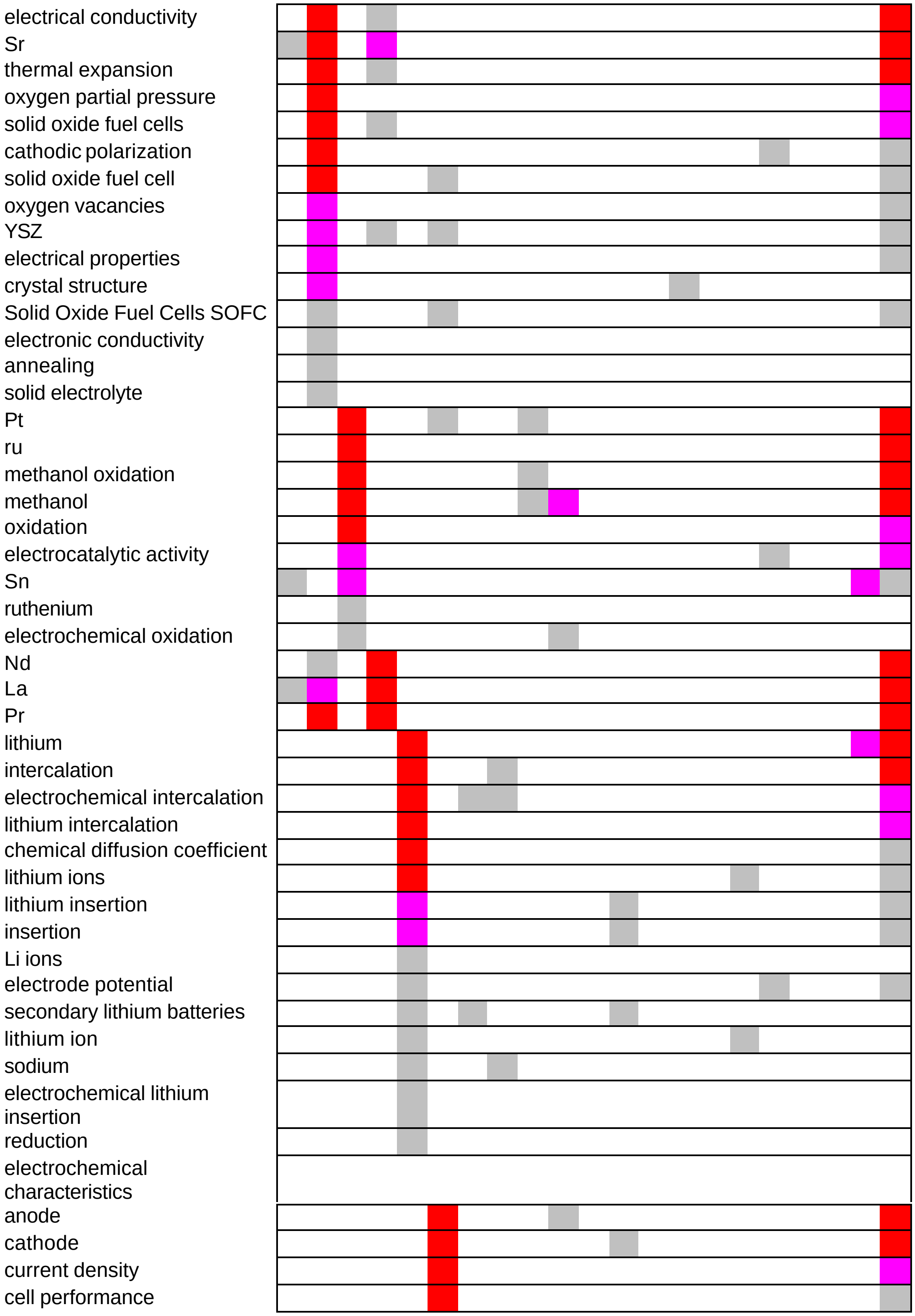
ECHEM SOURCES & STORAGE DEVICES (24,804)	BATTERIES (22,998)	Battery materials (7,850)	LITHIUM (LI, LIMN2O4, LICOO2, LITHIUM METAL, LINIO2), POLYMER, ALLOYS, CARBON (GRAPHITE), METAL, ACID, NICKEL (NI, NICKEL HYDROXIDE, NICKEL-CADMIUM), FILMS (THIN-FILM), HYDROGEN, HYDRIDE, ACTIVE MATERIALS, COMPOSITE , SALTS, MANGANESE, MAGNESIUM, SODIUM, IRON (IRON SULFIDE, SUPER-IRON, FES2,), COBALT, POWDER, ALUMINUM, ETHYLENE, HYDROXIDE, VANADIUM, CHLORIDE, GEL, CADMIUM, GLASS COPPER, SILVER, BISMUTH, ZINC, PLASTIC, LEAD SULFATE (PBSO4), MH ALLOY, SODIUM CHLORIDE,
		Properties and characteristics (4,643)	CAPACITY (DISCHARGE CAPACITY, SPECIFIC CAPACITY), DENSITY (ENERGY DENSITY), VOLTAGE, IMPEDANCE (INTERNAL IMPEDANCE), ONDUCTIVITY, PULSE, COULOMBIC EFFICIENCY, TEMPERATURES, RECHARGEABLE (RECHARGEABILITY), CYCLE LIFE (CYCLE PERFORMANCE), STABILITY (ELECTROCHEMICAL STABILITY), RESISTANCE, THIN, SEALED, HIGH ENERGY, AMORPHOUS, PORTABLE, DISCHARGED, AQUEOUS, LITHIATED, HIGH CAPACITY, CONDUCTIVE, MEMORY, IMPLANTABLE, LAMINATED, LIGHTWEIGHT

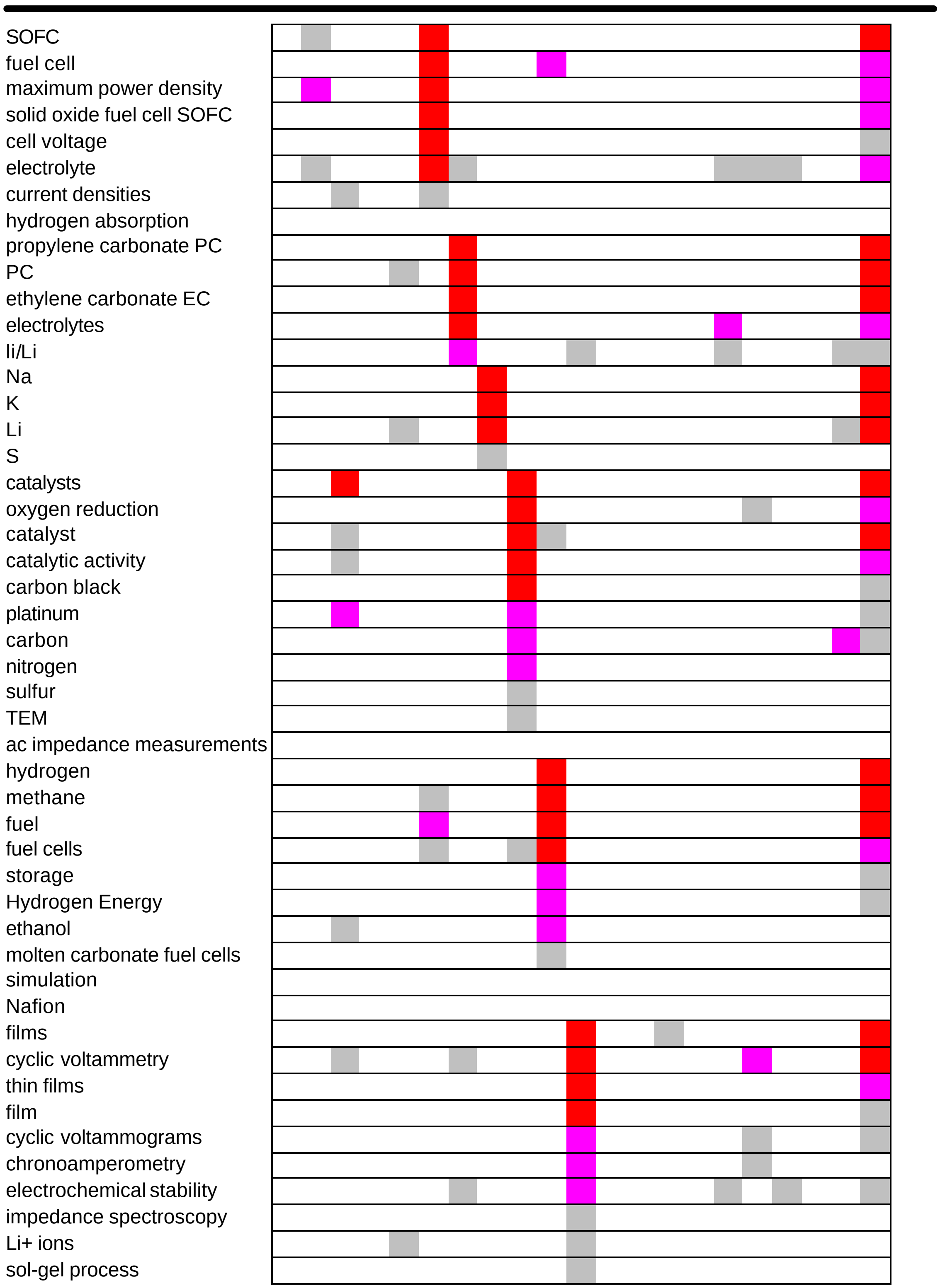
		Battery components (4,531)	ELECTRODES (POSITIVE ELECTRODES, NEGATIVE ELECTRODE), ELECTROLYTES (POLYMER ELECTROLYTES, LIQUID ELECTROLYTE, GEL POLYMER ELECTROLYTE, GELLED-ELECTROLYTE, FLOODED-ELECTROLYTE), CATHODE (COMPOSITE CATHODE), ANODE (CARBON ANODE), SEPARATORS, PLATES; STRAP, POSITIVE PLATE, COPPER CURRENT COLLECTOR
		Battery processes and phenomena (2,658)	DISCHARGE, CYCLING, CYCLES, INTERCALATION, CORROSION, CHARGING, CHARGE-DISCHARGE, OXIDATION, RECOVERY, REDOX FLOW, CAPACITY LOSS, FAILURE, SELF-DISCHARGE, OVERCHARGE, FLOODED, GRID CORROSION
		Battery types (2,195)	LITHIUM-ION, LEAD ACID, VRLA, RECHARGEABLE LITHIUM, LITHIUM POLYMER, ALKALINE, LITHIUM SECONDARY, NI MH, SOLID STATE, ORGANIC, PRIMARY, NICKEL-METAL HYDRIDE, SILVER-ZINC, VANADIUM REDOX, NICKEL CADMIUM, NICKEL-ZINC
		Battery applications (1,121)	ELECTRIC VEHICLES, GRIDS, ENERGY STORAGE, CIRCUIT, AUTOMOTIVE, POWER SOURCES, MOTOR, ELECTRONICS, AIRCRAFT, TRANSFORMER
	ECHEM CAPACITORS (1,806)	Capacitor properties and characteristics (604)	CAPACITANCE (SPECIFIC CAPACITANCE), DENSITY (ENERGY DENSITY), RESISTANCE, HIGH CONDUCTIVITY, ELECTROCHEMICAL, AQUEOUS, FATIGUE, FERROELECTRIC, CYCLIC, NONLINEAR, ELECTROLYTIC, POLARIZABLE, HIGH POWER DENSITY

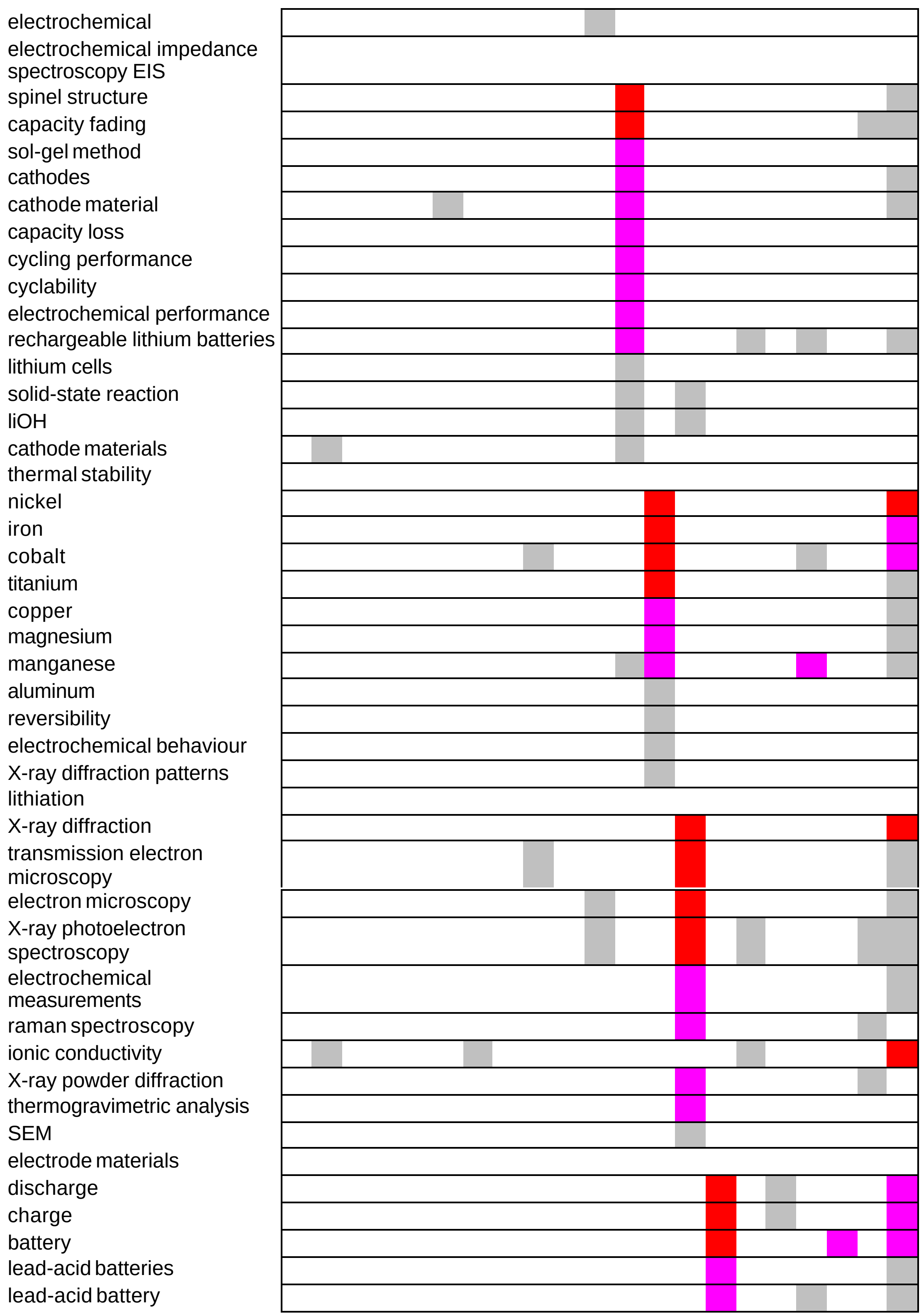
		Capacitor components (568)	ELECTROLYTE (GEL ELECTROLYTE, SOLID ELECTROLYTIC), ELECTRODES (POLARIZABLE ELECTRODES, BOTTOM ELECTRODE, CARBON ELECTRODES, COMPOSITE ELECTRODES, RUO2 ELECTRODES), FILMS (RUO2 THIN FILMS, NICKEL OXIDE FILMS), ACTIVATED CARBONS, CARBON CLOTH
		Capacitor materials (435)	CARBON (GLASSY CARBON, CARBON FIBER), RUO2 (RUO2 XH), GELS (SILICA GELS, AEROGELS, CARBON AEROGELS), THIN FILMS, TANTALUM, TETRAFLUOROBORATE, TETRAETHYLAMMONIUM, SI (SIO2), CLOTH, TETRAETHYLAMMONIUM TETRAFLUOROBORATE
		Capacitor types (199)	ELECTRIC DOUBLE-LAYER (EDLC), DIELECTRIC, CDL, PSEUDOCAPACITOR

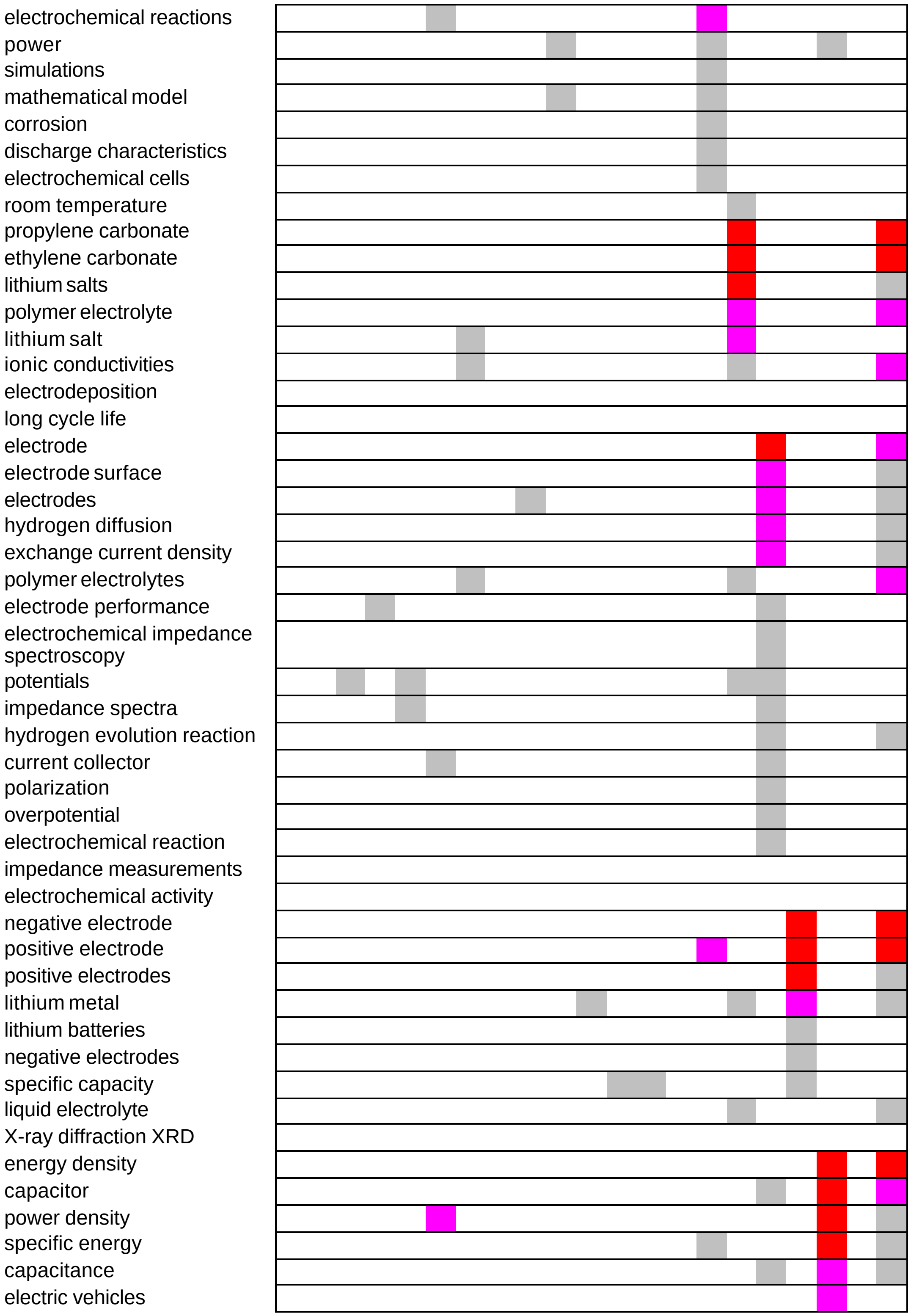
APPENDIX 7 – ABSTRACT PHRASE FACTOR MATRIX

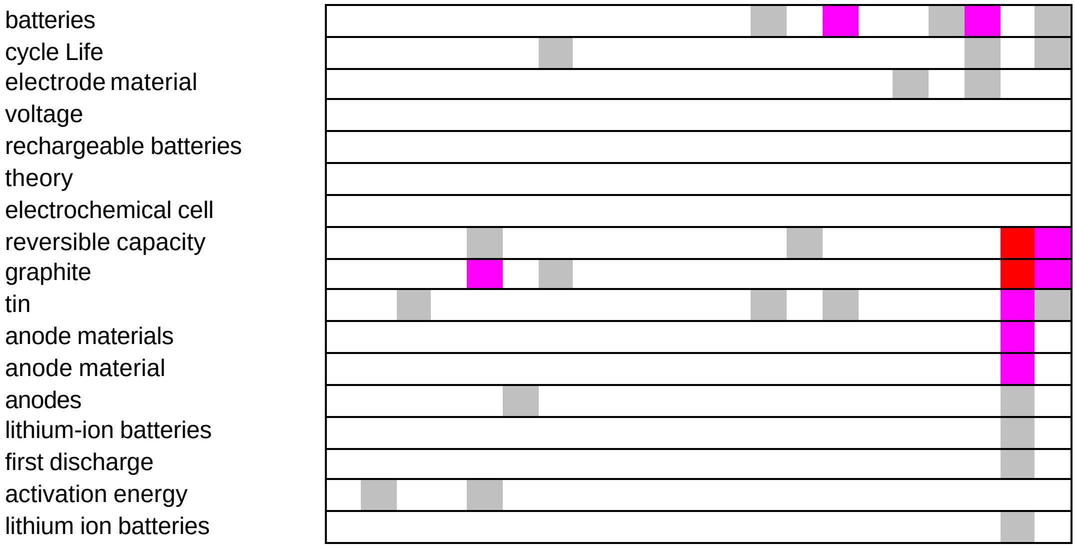



electrical conductivity
Sr
thermal expansion
oxygen partial pressure
solid oxide fuel cells
cathodic polarization
solid oxide fuel cell
oxygen vacancies
YSZ
electrical properties
crystal structure
Solid Oxide Fuel Cells SOFC
electronic conductivity
annealing
solid electrolyte
Pt
ru
methanol oxidation
methanol
oxidation
electrocatalytic activity
Sn
ruthenium
electrochemical oxidation
Nd
La
Pr
lithium
intercalation
electrochemical intercalation
lithium intercalation
chemical diffusion coefficient
lithium ions
lithium insertion
insertion
Li ions
electrode potential
secondary lithium batteries
lithium ion
sodium
electrochemical lithium insertion
reduction
electrochemical characteristics
anode
cathode
current density
cell performance

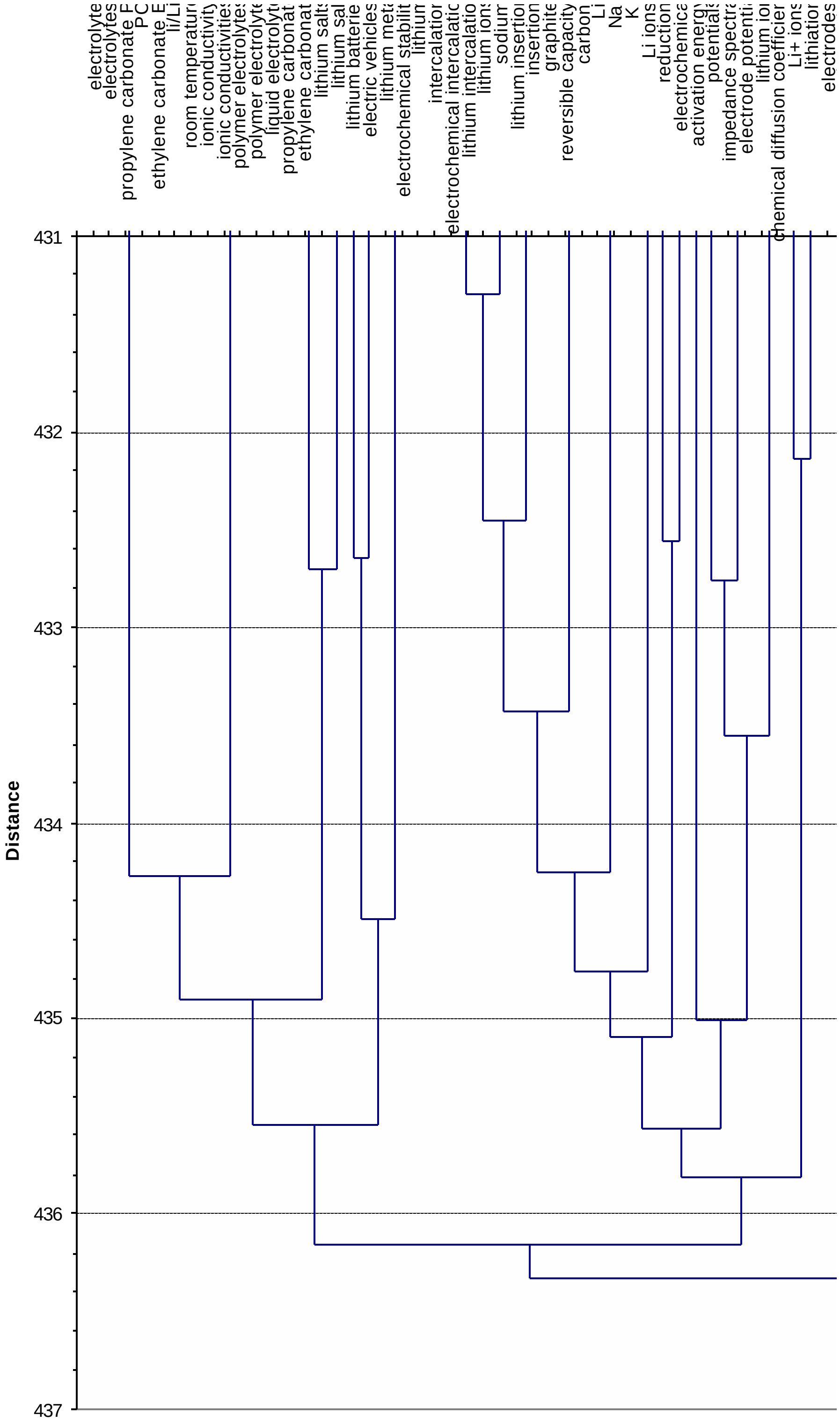

SOFC
fuel cell
maximum power density
solid oxide fuel cell SOFC
cell voltage
electrolyte
current densities
hydrogen absorption
propylene carbonate PC
PC
ethylene carbonate EC
electrolytes
li/Li
Na
K
Li
S
catalysts
oxygen reduction
catalyst
catalytic activity
carbon black
platinum
carbon
nitrogen
sulfur
TEM
ac impedance measurements
hydrogen
methane
fuel
fuel cells
storage
Hydrogen Energy
ethanol
molten carbonate fuel cells
simulation
Nafion
films
cyclic voltammetry
thin films
film
cyclic voltammograms
chronoamperometry
electrochemical stability
impedance spectroscopy
Li+ ions
sol-gel process


electrochemical
electrochemical impedance spectroscopy EIS
spinel structure
capacity fading
sol-gel method
cathodes
cathode material
capacity loss
cycling performance
cyclability
electrochemical performance
rechargeable lithium batteries
lithium cells
solid-state reaction
liOH
cathode materials
thermal stability
nickel
iron
cobalt
titanium
copper
magnesium
manganese
aluminum
reversibility
electrochemical behaviour
X-ray diffraction patterns
lithiation
X-ray diffraction
transmission electron microscopy
electron microscopy
X-ray photoelectron spectroscopy
electrochemical measurements
raman spectroscopy
ionic conductivity
X-ray powder diffraction
thermogravimetric analysis
SEM
electrode materials
discharge
charge
battery
lead-acid batteries
lead-acid battery

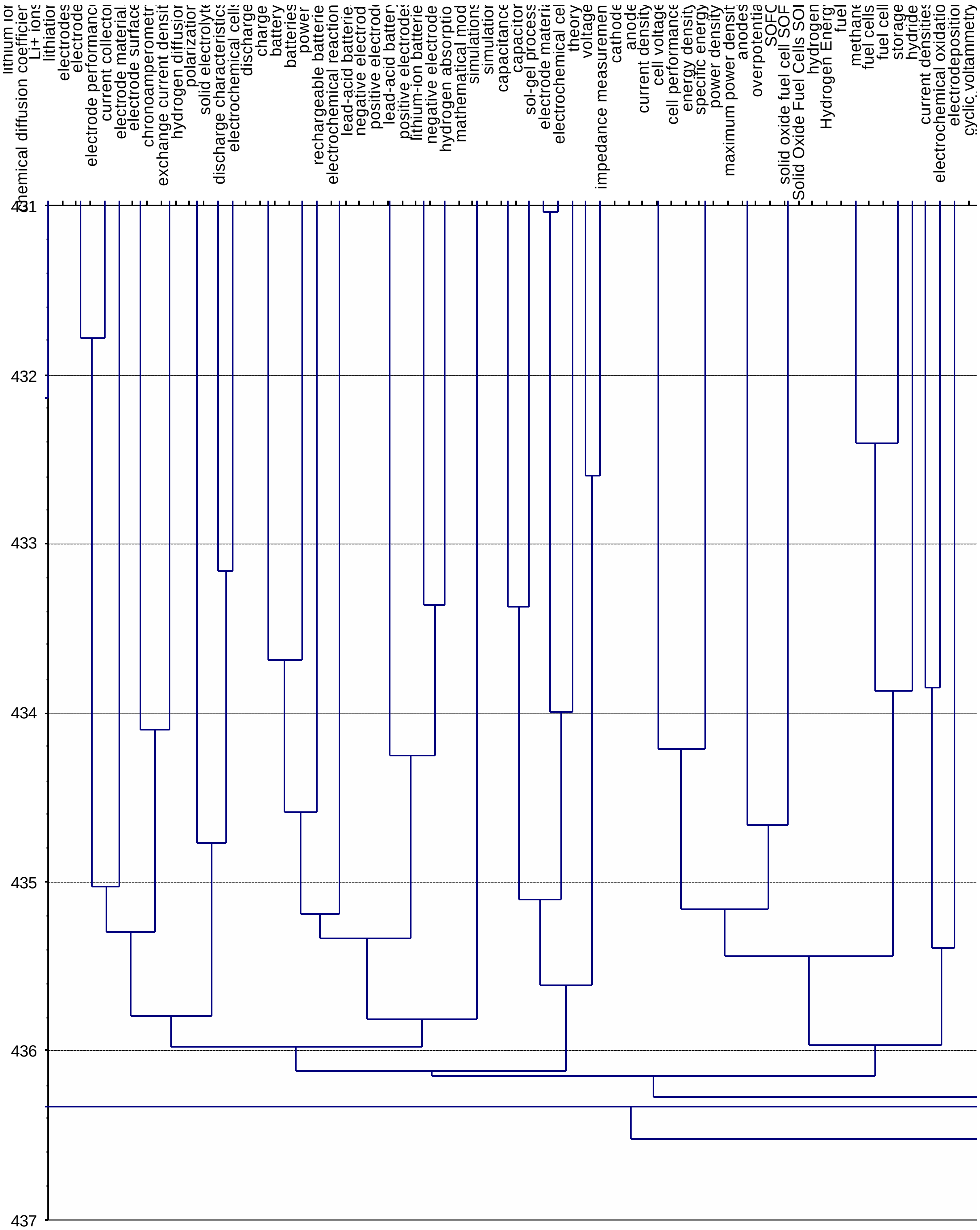

electrochemical reactions
power
simulations
mathematical model
corrosion
discharge characteristics
electrochemical cells
room temperature
propylene carbonate
ethylene carbonate
lithium salts
polymer electrolyte
lithium salt
ionic conductivities
electrodeposition
long cycle life
electrode
electrode surface
electrodes
hydrogen diffusion
exchange current density
polymer electrolytes
electrode performance
electrochemical impedance spectroscopy
potentials
impedance spectra
hydrogen evolution reaction
current collector
polarization
overpotential
electrochemical reaction
impedance measurements
electrochemical activity
negative electrode
positive electrode
positive electrodes
lithium metal
lithium batteries
negative electrodes
specific capacity
liquid electrolyte
X-ray diffraction XRD
energy density
capacitor
power density
specific energy
capacitance
electric vehicles

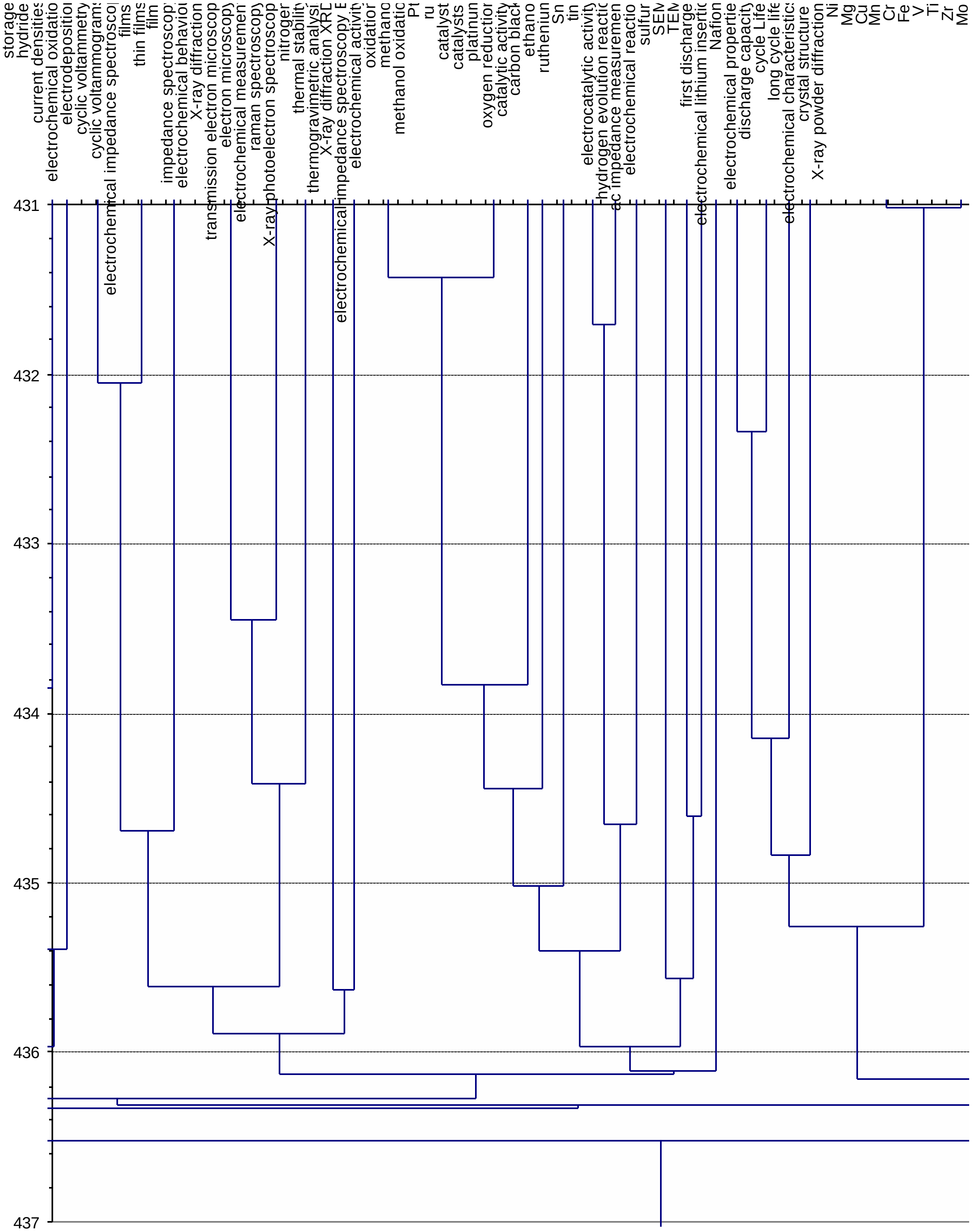

batteries
cycle Life
electrode material
voltage
rechargeable batteries
theory
electrochemical cell
reversible capacity
graphite
tin
anode materials
anode material
anodes
lithium-ion batteries
first discharge
activation energy
lithium ion batteries

APPENDIX 8 – ABSTRACT PHRASE CLUSTERING DENDOGRAM



APPENDIX 8 – ABSTRACT PHRASE CLUSTERING DENDOGRAM (CONTINUED)




storage
hydride
current densities
electrochemical oxidation
electrodeposition
cyclic voltammetry
cyclic voltammograms
electrochemical impedance spectroscopy
films
thin films
film
impedance spectroscopy
electrochemical behavior
X-ray diffraction
transmission electron microscopy
electron microscopy
electrochemical measurements
raman spectroscopy
X-ray photoelectron spectroscopy
nitrogen
thermal stability
thermogravimetric analysis
X-ray diffraction XRD
electrochemical impedance spectroscopy
electrochemical activity
oxidation
methanol
methanol oxidation
Pt
ru
catalyst
catalysts
platinum
oxygen reduction
catalytic activity
carbon black
ethanol
ruthenium
Sn
tin
electrocatalytic activity
hydrogen evolution reaction
ac impedance measurements
electrochemical reaction
sulfur
SEM
TEM
first discharge
electrochemical lithium insertion
Nafion
electrochemical properties
discharge capacity
cycle Life
long cycle life
electrochemical characteristics
crystal structure
X-ray powder diffraction
Ni
Mg
Cu
Mn
Cr
Fe
V
Ti
Zr
Mo
431
432
433
434
435
436
437

APPENDIX 8 – ABSTRACT PHRASE CLUSTERING DENDOGRAM (CONTINUED)